

Copyright

by

Daniel Urieli

2015

The Dissertation Committee for Daniel Urieli  
certifies that this is the approved version of the following dissertation:

## **Autonomous Trading in Modern Electricity Markets**

Committee:

---

Peter Stone, Supervisor

---

Raymond Mooney

---

Pradeep Ravikumar

---

Ross Baldick

---

Zico Kolter

# **Autonomous Trading in Modern Electricity Markets**

by

**Daniel Urieli, B.S., M.S.**

## **Dissertation**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas at Austin**

December 2015

To my family

# Acknowledgments

I would like to thank the people who have made my Ph.D. possible, and to those who have made it an enriching, exciting, enjoyable, and memorable experience.

It has been a great honor to work with Prof. Peter Stone – a world class AI researcher – as my advisor. Peter, I would like to express here my deep gratitude to you. You have been the best advisor I could have hoped for. Your outstanding professional capabilities and your rare personal qualities have made working with you a unique experience I will always cherish. Working with you has made me realize how the impossible can become possible. You have created a unique atmosphere in our group: of friendship, trust, honesty, support, and of striving to win. By your deep and broad understanding of AI, and by providing the right guidance, you have been a mentor and a colleague in an exciting journey into the world of AI research. I deeply thank you for that.

I would like to thank my committee members: Ross Baldick, Raymond J. Mooney, Pradeep Ravikumar, and J. Zico Kolter. To Ross, for helping me to get exposed to the world of power markets, for your helpful advice along the way, and for being so responsive, professional and kind. To Ray, Pradeep, and Zico, for your sharp insights and excellent advice, which helped shaping my dissertation.

I would like to thank my fellow members of the Learning Agents Research Group (LARG), who are talented researchers and good friends. Shivaram Kalyanakrishnan has been amazingly helpful and provided valuable advice during my early

days in grad school. Noa Agmon has been a brilliant collaborator of mine and a good friend. David Pardoe, the consistent champion of the Trading Agent Competitions (TAC), kindly introduced me to the TAC domain, which became the focus of my dissertation. Exchanges with members of the group contributed directly and indirectly to this dissertation, including with the current Ph.D. student members of our group: Patrick MacAlpine, Matthew Hausknecht, Katie Genter, Piyush Khandelwal, Jacob Menashe, Elad Liebman, Sanmit Narvekar, and Josiah Hanna; the current and former Post-Doc members of our group: Michael Albert, Jivko Sinapov, Shiqi Zhang, Matteo Leonetti, Tsz-Chiu Au, Michael Quinlan, Tobias Jung; and former Ph.D. students of our group: Matt Taylor, Greg Kuhlmann, Nick Jong, Juhyun Lee, Doran Chakraborty, Brad Knox, Todd Hester, Yinon Bentor, and Samuel Barrett. I am deeply thankful for their contribution to my work.

I would like to thank all the developers and participants of the Power Trading Agent Competition (Power TAC) for creating and cultivating a vibrant community focused on applying modern AI to the benefit of sustainable environment, through a unique combination of an exciting competition and a rich research platform.

I have been fortunate to be a recipient of the “NSF IGERT - Sustainable Grid Integration of Distributed and Renewable Resources” fellowship, which included being a member of a selective interdisciplinary group of energy research students. My membership in this group has broadened my view on sustainable energy. I would like to thank the group members for their valuable insights, and especially to Dr. Tom Edgar who runs the program, to Risa Hartman and Darlene Yanez for their incredible support along the way, and to NSF IGERT for their support.

UT Austin’s Computer Science Department has provided an amazing working environment. I would like to thank the technical staff – the Gripe team – for providing and maintaining computing infrastructure of the highest quality; to the excellent administrative staff, and especially to Stacy Miller and Lydia Griffith.

I would like to dedicate a special thanks to Katherine Utz, whose professional responses to my inquiries as a prospective student made me feel that UT Austin might be a great home for my graduate studies. I was right.

I would also like to thank Tel Aviv University for providing me with a rigorous computer science and mathematics education which maintained standards of the highest level. I am especially thankful to my M.Sc. advisor, Prof. Yossi Matias, and to Prof. Shmuel (Mooly) Sagiv and Prof. Yossi Azar, who are brilliant educators.

I wish to thank my family, and first and foremost to my parents, Elina and Israel Urieli. Mom, Dad: words can barely convey my gratitude to you. You have always provided me with an intellectually stimulating environment. Your trust and confidence in me have made me believe I could succeed at any undertaking; the thoughtful way in which you have educated me has provided me the tools to do so. I am endlessly grateful for everything you have given me. To my dear sisters Noa and Ayelet: I am forever grateful for your immeasurable love and support throughout the years. To my sisters' husbands Gil and Ron: thank you for stimulating discussions which have directly impacted my research. To my parents in law, Ron and Sari: thank you for your love and support all along the way. To the rest of my family: thank you for all your love and caring.

I would like to end with two special dedications. To my sweet little daughters Moriah and Nogah: thank you for bringing so much joy to our home, for inspiring me, and for making me such a proud dad. Most of all, to my dear beautiful wife Adi: thank you for believing in me, for being a source of confidence and happiness, for your trust and support, and for your endless love; I love you!

DANIEL URIELI

*The University of Texas at Austin*

*December 2015*

# Autonomous Trading in Modern Electricity Markets

Daniel Urieli, Ph.D.

The University of Texas at Austin, 2015

Supervisor: Peter Stone

The smart grid is an electricity grid augmented with digital technologies that automate the management of electricity delivery. The smart grid is envisioned to be a main enabler of sustainable, clean, efficient, reliable, and secure energy supply. One of the milestones in the smart grid vision will be programs for customers to participate in electricity markets through *demand-side management* and distributed generation; electricity markets will (directly or indirectly) incentivize customers to adapt their demand to supply conditions, which in turn will help to utilize intermittent energy resources such as from solar and wind, and to reduce peak-demand.

Since wholesale electricity markets are not designed for individual participation, *retail brokers* could represent customer populations in the wholesale market, and make profit while contributing to the electricity grid's stability and reducing customer costs. A retail broker will need to operate continually and make real-time decisions in a complex, dynamic environment. Therefore, it will benefit from employing an *autonomous broker agent*. With this motivation in mind, this dissertation



makes five main contributions to the areas of artificial intelligence, smart grids, and electricity markets.

First, this dissertation formalizes the problem of autonomous trading by a retail broker in modern electricity markets. Since the trading problem is intractable to solve exactly, this formalization provides a guideline for approximate solutions.

Second, this dissertation introduces a general algorithm for autonomous trading in modern electricity markets, named LATTE (Lookahead-policy for Autonomous Time-constrained Trading of Electricity). LATTE is a general framework that can be instantiated in different ways that tailor it to specific setups.

Third, this dissertation contributes fully implemented and operational autonomous broker agents, each using a different instantiation of LATTE. These agents were successful in international competitions and controlled experiments and can serve as benchmarks for future research in this domain. Detailed descriptions of the agents' behaviors as well as their source code are included in this dissertation.

Fourth, this dissertation contributes extensive empirical analysis which validates the effectiveness of LATTE in different competition levels under a variety of environmental conditions, shedding light on the main reasons for its success by examining the importance of its constituent components.

Fifth, this dissertation examines the impact of Time-Of-Use (TOU) tariffs in competitive electricity markets through empirical analysis. Time-Of-Use tariffs are proposed for demand-side management both in the literature and in the real-world.

The success of the different instantiations of LATTE demonstrates its generality in the context of electricity markets. Ultimately, this dissertation demonstrates that an autonomous broker can act effectively in modern electricity markets by executing an efficient lookahead policy that optimizes its predicted utility, and by doing so the broker can benefit itself, its customers, and the economy.

# Contents

<b>Acknowledgments</b>	<b>v</b>
<b>Abstract</b>	<b>viii</b>
<b>List of Tables</b>	<b>xiv</b>
<b>List of Figures</b>	<b>xv</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Objectives and Approach . . . . .	3
1.1.1 Dissertation Research Question . . . . .	3
1.1.2 Approach . . . . .	4
1.2 Contributions . . . . .	5
1.3 Dissertation Overview . . . . .	6
<b>Chapter 2 Substrate Domain: The Power TAC Simulator</b>	<b>9</b>
2.1 Power TAC Overview . . . . .	10
2.1.1 The Power TAC Competition . . . . .	10
2.1.2 The Power TAC Simulation Environment: An Overview . . . .	10
2.2 Broker Interaction with the Environment . . . . .	15
2.2.1 Broker Actions . . . . .	15
2.2.2 Environment Response to Broker Actions . . . . .	18

2.2.3	Broker Sensing . . . . .	23
2.3	Power TAC Modeling Assumptions . . . . .	25
2.4	Chapter Summary . . . . .	28
<b>Chapter 3 The Broker’s Power Trading Problem: Formalization</b>		<b>29</b>
3.1	The Power Trading Problem’s Temporal Structure . . . . .	30
3.2	Power Trading as a Markov Decision Process . . . . .	30
3.2.1	Dimensionality of the State and Action Spaces . . . . .	36
3.3	Lookahead Policies as Approximate Solutions . . . . .	38
3.4	Chapter Summary . . . . .	40
<b>Chapter 4 The LATTE Algorithm</b>		<b>41</b>
4.1	The Design Principles of LATTE . . . . .	42
4.2	The LATTE Algorithm . . . . .	45
4.3	Chapter Summary . . . . .	49
<b>Chapter 5 TacTex-13: A Champion Adaptive Power Trading Agent</b>		<b>51</b>
5.1	TacTex-13 Description . . . . .	52
5.1.1	TacTex-13’s Instantiation of LATTE’s Phase 1 . . . . .	52
5.1.2	TacTex-13’s Instantiation of LATTE’s Phase 2 . . . . .	57
5.2	Results . . . . .	61
5.2.1	Competition Results: Power TAC 2013 Finals Analysis . . . .	62
5.2.2	Controlled Experiments . . . . .	64
5.3	Chapter Summary . . . . .	67
<b>Chapter 6 TacTex-15: A Winning Power Trading Agent</b>		<b>68</b>
6.1	TacTex-15 Agent Description . . . . .	69
6.1.1	Instantiating PredictTariffEffects: a New Demand-Predictor . .	69
6.1.2	Instantiating PredictWholesalePrice: a New Cost-Predictor . .	70

6.1.3	Instantiating WholesaleBiddingPolicy . . . . .	73
6.2	Results . . . . .	74
6.2.1	Competition Results: Power TAC 2015 Finals Analysis . . . .	74
6.2.2	Controlled Experiments . . . . .	79
6.3	Chapter Summary . . . . .	87
<b>Chapter 7</b>	<b>Autonomous Trading using Time-Of-Use Tariffs</b>	<b>90</b>
7.1	Background . . . . .	92
7.1.1	Motivation: DSM and TOU Tariffs . . . . .	92
7.1.2	DSM in Power TAC . . . . .	94
7.2	Our Contribution to DSM in Power TAC . . . . .	95
7.3	Challenges of Using TOU Tariffs in Competitive Markets . . . . .	97
7.4	LATTE-TOU . . . . .	98
7.4.1	LATTE as a Black-Box Optimization . . . . .	98
7.4.2	LATTE-TOU . . . . .	100
7.5	Results . . . . .	102
7.5.1	LATTE-TOU's Impact on the Broker and the Economy . . . .	103
7.5.2	Robustness of TOU to Prediction Errors . . . . .	106
7.6	Chapter Summary . . . . .	108
<b>Chapter 8</b>	<b>Related Work</b>	<b>109</b>
8.1	Smart Grid and Electricity Markets . . . . .	110
8.1.1	Time-Of-Use Tariffs and Demand-Side Management . . . . .	111
8.2	MAS for the Smart Grid . . . . .	113
8.3	Agent-Based Power Market Simulations . . . . .	115
8.4	Autonomous Trading Agents and TAC . . . . .	116
8.5	Power TAC Broker Agents . . . . .	118
8.5.1	Early Power TAC Brokers . . . . .	118

8.5.2	Power TAC 2013 agents . . . . .	120
8.5.3	Power TAC 2014 agents . . . . .	122
8.5.4	Power TAC 2015 . . . . .	123
8.6	Chapter Summary . . . . .	123
<b>Chapter 9 Conclusions and Future Work</b>		<b>125</b>
9.1	Contributions . . . . .	125
9.2	Future Work . . . . .	127
9.2.1	Directions for Extending LATTE within Power TAC . . . . .	128
9.2.2	Directions for Extending Power TAC . . . . .	132
9.2.3	Directions for Extending LATTE to Real Markets . . . . .	135
9.3	Concluding Remarks . . . . .	137
<b>Appendix A TacTex Agents Source Code, Binaries, and Resources</b>		<b>138</b>
<b>Appendix B Power TAC Game Parameters</b>		<b>139</b>
<b>Bibliography</b>		<b>140</b>

# List of Tables

2.1	Tariff market actions. . . . .	17
2.2	Retail tariffs used in this dissertation. . . . .	18
2.3	Wholesale market actions. . . . .	19
3.1	Variables used for the power trading MDP’s main components. . . .	32
3.2	Variables used in the power trading MDP’s state definition. . . . .	33
3.3	Variables used in the power trading MDP’s reward definition. . . . .	34
5.1	Results of the Power TAC 2013 finals . . . . .	62
5.2	Round-Robin ablation analysis. . . . .	65
5.3	Ablation analysis using 3 finalist broker agents. . . . .	66
6.1	Power TAC 2015 finals results. . . . .	76
6.2	Power TAC 2015 post-finals demo competition results. . . . .	76
6.3	Payoff matrix of two wholesale-bidding strategies in 2-agent games. .	87
7.1	Comparing TacTex-TOU with two baselines: TOUNaive and FixedRate	105
7.2	TacTex-TOU: self-play . . . . .	105
7.3	Ablation analysis: erratic-predictions . . . . .	107
B.1	Power TAC Game Parameters . . . . .	139

# List of Figures

1.1	Guide for reading individual chapters. . . . .	8
2.1	Structure of the Power TAC simulation environment . . . . .	11
2.2	Broker interactions with the simulation environment. . . . .	16
2.3	Double-Auction Clearing Example. . . . .	22
3.1	Temporal structure of the power trading problem. . . . .	31
4.1	The LATTE algorithm. . . . .	46
5.1	Power TAC 2013 finals: average income and costs . . . . .	63
6.1	2015 competition analysis: average profit, revenue and costs. . . . .	78
6.2	Controlled experiments: TacTex-15 vs. Power TAC 2015 finalists . . .	80
6.3	Ablation analysis for 2-6 broker games. . . . .	82
6.4	Performance dependence on demand-predictor in 3-agent games. . . .	83
6.5	Demand-predictor ablation: revenue and cost components. . . . .	85
6.6	Demand-predictor ablation: broker behavior. . . . .	86
6.7	Demand-predictor ablation: customers and profits. . . . .	86
6.8	Cost-predictor ablation in presence of abruptly changing market-costs.	88
7.1	Generation-cost curves of ERCOT, PJM, CAISO. . . . .	93

7.2	Consumption flattening. . . . .	107
8.1	Dissertation contribution areas. . . . .	110



# Chapter 1

## Introduction

To ensure sustainable existence for society, a transition to sustainable energy consumption is necessary. In 2003, the U.S. Department of Energy published a report named “Grid 2030: A National Vision For Electricity’s Second 100 Years” [114] which lays out a vision and an action plan for upgrading the traditional electricity grid into a *smart grid*. The smart grid is expected to be a major step towards sustainable, clean, efficient, reliable, and secure energy supply. Current deployments of smart grid technologies [108] and planned investments in the smart grid show that there is high-interest around the world to make the smart grid vision a reality. In the US, the required investments between 2011 and 2031 are estimated to total close to 500 billion dollars, and the benefit is estimated to be 1.2-2 trillion dollars [17]. In the European Union, the required investments between 2011 and 2050 are estimated to total about 2 trillion Euros [18].

One of the milestones described in the Grid 2030 vision is “Programs for customer participation in power markets through demand-side management and distributed generation”. Demand-side management refers to adapting customers’ electricity demand to supply conditions, and may be implemented using new power

markets (aka electricity markets)<sup>1</sup> that financially motivate desirable demand patterns. Since wholesale power markets are not designed for individual customer participation [49], retail brokers can represent large customer populations. These retail brokers can aggregate and predict customers' demand and production and participate in the wholesale market on behalf of their customers. By representing customer populations, brokers can make profit while reducing their customers' costs and contributing to grid stability [44, 47, 45]. Grid stability depends on having a perfect balance between supply and demand at all times. Brokers who are financially incentivized to maintain supply-demand balance in their portfolios can financially incentivize their customers to adapt their demand to supply conditions, thus contributing to demand-side management, and to the above Grid 2030 milestone.

A retail broker acts in multiple markets in parallel. In the retail market, it designs tariff contracts that attract consumers and distributed producers (such as rooftop solar and wind turbines). In the wholesale market, it bids for future energy contracts. At all times the broker must maintain supply-demand balance in its portfolio, potentially by affecting its customers' demand through pricing incentives and consumption curtailments.

Operating profitably as a retail broker is a challenging problem. A broker needs to continually select among a large set of actions, under real-time constraints, while incorporating large amounts of information and complex calculations into its decision process, so that its long term profit is maximized in a competitive, dynamic, and stochastic environment. Due to these problem characteristics, it can be beneficial to employ an *autonomous broker agent* to decide and act on behalf of the retail broker.

The idea of using autonomous electricity broker agents has been proposed by Ketter et al. [44] and motivates the line of research in this dissertation. Russell and

---

<sup>1</sup>Throughout this dissertation we use the terms energy/electricity/power interchangeably.

Norvig categorized agent task environments along a number of dimensions [93]. The broker’s electricity trading task environment would be categorized using the most complex option along each of these dimensions, being *partially-observable*, *multi-agent*, *competitive*, *stochastic*, *sequential*, *dynamic*, *continuous*, *initially-unknown*. On top of being continuous (having continuous world state and actions), this task environment is also *high-dimensional*, having high-dimensional state and action representations.

Due to the complexity of the broker’s electricity trading problem, a first observation that can be made is that designing an autonomous broker that acts optimally would be an impossible task. Thus, a primary research goal of this dissertation is designing and investigating autonomous electricity trading strategies that approximate the optimal strategy and perform well empirically.

## 1.1 Objectives and Approach

### 1.1.1 Dissertation Research Question

The principal question addressed in this dissertation is:

How should an autonomous broker agent act to maximize its utility by trading in time-constrained, modern electricity markets?

This dissertation advances towards answering this question by contributing a general decision-making framework for an autonomous broker agent in smart grid environments where:

- A broker’s goal is maximizing its long term utility (profit).
- A broker decides on actions in real-time.
- A broker publishes tariff contracts to which retail customers can subscribe.
- A broker bids to buy and sell electricity in a day-ahead wholesale market.

- Supply-demand imbalance incurs (positive or negative) payments on brokers.

### 1.1.2 Approach

Our general approach to answering the dissertation research question is to proceed from theory to practice using a series of approximations: we start by formalizing the complete electricity trading problem and characterizing its optimal solution; we then use the problem formalization to guide us through a series of approximations into practical solutions; we then create complete autonomous broker agents that implement these practical solutions; finally, we test these brokers extensively in a complex, realistic electricity trading environment, under a variety of conditions and competition levels.

Electricity markets are going through a major transition from traditional, regulated monopolies into deregulated, competitive markets [35]. While in principle, deregulation can increase efficiency, in practice, the California energy crisis (2001) has demonstrated the high-costs of failure due to flawed deregulation [99, 7], and the importance of testing new market structures in simulation before deploying them. This is the focus of the Power Trading Agent Competition (Power TAC) [44, 47, 45], which we use throughout this dissertation as a substrate domain for our research.

Power TAC is a realistic, detailed platform for modeling and testing competitive retail power market designs and related automation technologies. Power TAC simulates a future smart grid environment with about 57,000 customers (about 50,000 consumers and 7,000 renewable producers), smart-metering, autonomous agents acting on behalf of customers and retailers, and realistic market designs: the wholesale market represents a traditional energy exchange, such as Nord Pool or EEX, and the retail market is similar to ERCOT's<sup>2</sup>. Power TAC's customers are simulated using state-of-the-art customer models [87]. In Power TAC's envi-

---

<sup>2</sup>See [www.nordpoolspot.gov](http://www.nordpoolspot.gov), [www.eex.com](http://www.eex.com), [www.ercot.com](http://www.ercot.com)

ronment, autonomous broker agents compete with each other to make profits by trading in retail, wholesale, and balancing markets (the balancing market provides a simplified version of the reserve and regulating capacity markets and associated controls normally operated by an ISO/TSO organization [46]).

Research results from Power TAC may help policy makers create mechanisms that produce the intended incentives for energy producers and consumers. They are also expected to help to develop and validate intelligent automation technologies that can support effective management of participants in electricity markets. Due to the high-level of detail and realism of Power TAC, we believe it can be viewed as a reasonable substrate domain for studying general electricity market conditions.

## 1.2 Contributions

This dissertation makes five distinct contributions at the intersection between artificial intelligence (AI), smart grids, and electricity markets:

**Problem Formalization** A formal specification of the problem of autonomous trading by a retail broker in modern electricity markets. This formalization is suitable when brokers trade in the retail market by publishing tariff contracts, trade in the wholesale market by bidding for future contracts, and where supply-demand imbalance results in (positive or negative) payments by the broker. This problem is formalized as a Markov Decision Process, which due to its complexity is intractable to solve exactly.

**The LATTE Algorithm** A general algorithm that approximates the solution to the autonomous broker trading problem. LATTE stands for Lookahead-policy for Autonomous Time-constrained Trading of Electricity. LATTE is a general framework that can be instantiated in different ways that tailor it to specific setups.

**Complete Agents using Instantiations of LATTE** Fully implemented and operational autonomous broker agents, each using a different instantiation of LATTE. These agents were successful in competitions and controlled experiments and can serve as benchmarks for future research in the power trading domain. Detailed descriptions of the agents’ behaviors as well as their source code are included in this dissertation.

**Extensive Empirical Analysis** Extensive empirical analysis validates the effectiveness of LATTE, shedding light on the main reasons for its success by examining the importance of its constituent components.

**Impact of Time-Of-Use Tariffs in Competitive Markets** A gradient-ascent algorithm for optimizing Time-Of-Use tariffs by an autonomous broker in competitive markets (as a part of one of the instantiations of LATTE), and an empirical analysis of the impact of Time-Of-Use tariffs in competitive markets, on an autonomous broker and on the market. Time-Of-Use tariffs are a main method proposed for demand-side management both in the literature and in real-markets.

From the perspective of AI, these contributions are a demonstration of effective sequential decision making in a complex *partially-observable, multiagent, competitive, stochastic, sequential, dynamic, continuous, high-dimensional, initially-unknown* domain.

## 1.3 Dissertation Overview

The remainder of the dissertation is organized as follows.

**Chapter 2** introduces the substrate domain used in this dissertation: The Power Trading Agent Competition (Power TAC) simulation environment. Power TAC’s

simulation environment is a detailed, realistic electricity markets simulator that is expected to help policy makers to create appropriate market mechanisms, and to help to develop and validate intelligent automation technologies for electricity markets.

**Chapter 3** formalizes the problem faced by an autonomous electricity trading broker agent as a Markov Decision Process, which is impossible to solve exactly. Therefore, this chapter provides guidelines for effectively approximating its solution.

**Chapter 4** introduces the LATTE algorithm, which is a general algorithm for autonomous trading in modern electricity markets. LATTE stands for Lookahead-policy for Autonomous Time-constrained Trading of Electricity.

**Chapter 5** introduces the first instantiation of LATTE, used by the TacTex-13 agent, which won 1st place in the 2013 Power TAC finals. The chapter details TacTex-13's constituent components and analyzes their contribution to TacTex-13's success.

**Chapter 6** introduces a second instantiation of LATTE, used by the TacTex-15 agent, which achieved top performance in the Power TAC 2015 finals. The chapter analyzes the performance of TacTex-15 in competitions and controlled experiments, and the contributions of TacTex-15's constituent components to its performance.

**Chapter 7** enables and enhances Power TAC customers' demand-shifting capabilities. This chapter then introduces a third instantiation of LATTE that uses Time-Of-Use tariffs. It investigates how a broker should select such tariffs, and what their impacts are on the broker and on the market.

**Chapter 8** surveys related work in the electricity markets literature and in the

artificial intelligence and autonomous agents literature.

**Chapter 9** summarizes the contributions of this dissertation, and outlines promising directions for future work.

For readers who may wish to read chapters out of order, Figure 1.1 presents a diagram specifying the relations between the different chapters. For example, a reader that may wish to read Chapter 7, will need to first read Chapter 2, then Chapter 3, then Chapter 4 before reading Chapter 7. This reader will find Chapter 6 useful but not necessary to understand Chapter 7, and Chapter 5 useful but not necessary to understand Chapter 6.

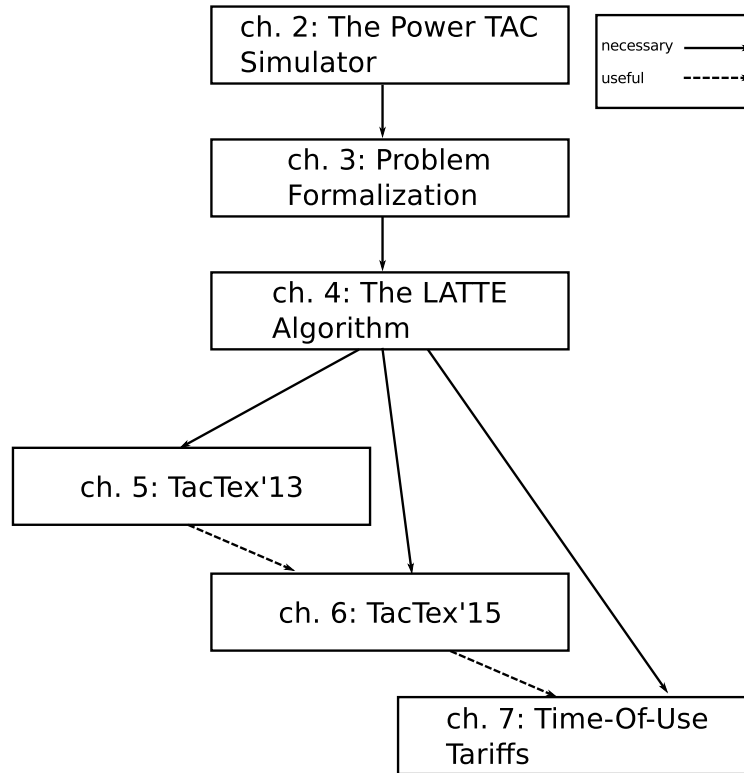


Figure 1.1: Guide for reading individual chapters: a solid arrow is drawn from a chapter that is necessary as a background for a following chapter, and a dashed arrow is drawn from a chapter that is useful as a background for a following chapter.



## Chapter 2

# Substrate Domain: The Power TAC Simulator

This chapter describes the substrate domain used in this dissertation: the Power Trading Agent Competition (Power TAC) simulation environment. The full details of the Power TAC game are specified in the official game description [46]. This chapter focuses on the aspects of the game that are most essential for understanding the rest of the dissertation. In Power TAC, autonomous brokers compete with each other to make profit by acting in multiple electricity markets in a simulated smart grid environment. Section 2.1 overviews the competition and the simulation environment. Section 2.2 details the actions available to brokers, the environment’s response to these actions, and the environment sensors available to brokers. Section 2.3 lists Power TAC’s modeling assumptions. The Power TAC simulation environment is open-source, and can be downloaded from GitHub.<sup>1</sup> A full specification of the Power TAC game can be found in the Power TAC game description [46].

---

<sup>1</sup><https://github.com/powertac>

## 2.1 Power TAC Overview

This section overviews the Power TAC competition and simulation environment.

### 2.1.1 The Power TAC Competition

Power TAC is an annual competition in which the competitors are autonomous brokers programmed by teams from around the world. The competition includes hundreds of games and takes several days to complete. In a game, the Power TAC simulator runs on a central server, while competing brokers run remotely and communicate with the server through the internet. Each broker receives partial state information from the server, and responds by communicating the actions it takes. The competition includes different game sizes, ranging from a small to large number of competitors. After the competition, participants release their broker binaries, which allows for running controlled experiments against the state of the art brokers.

### 2.1.2 The Power TAC Simulation Environment: An Overview

This section overviews the Power TAC simulation environment. Power TAC uses a rich, detailed power markets simulator, modeling a smart grid environment of a medium-sized city. Figure 2.1 shows the structure of the Power TAC simulation environment. Electricity is generated both by traditional generation companies which generate on demand, and by distributed renewable producers, who are also retail customers, which generate based on weather conditions. Electricity is consumed by a variety of commercial and residential consumer customers. Power TAC has more than 57,000 simulated customers (50,000 consumers and 7,000 renewable producers). Power TAC’s customers are *autonomous agents* that optimize the electricity-costs and convenience of their human owners [87]. Customers represent commercial and residential buildings, hospitals, distributed solar panels, wind farms, storage facilities and electric vehicles. Consumers and producers consume/produce using

time-series generators constructed from real-world data, according to weather and calendar factors. Weather conditions are determined by files containing real-world weather data, recorded in different geographical zones at different times of year.

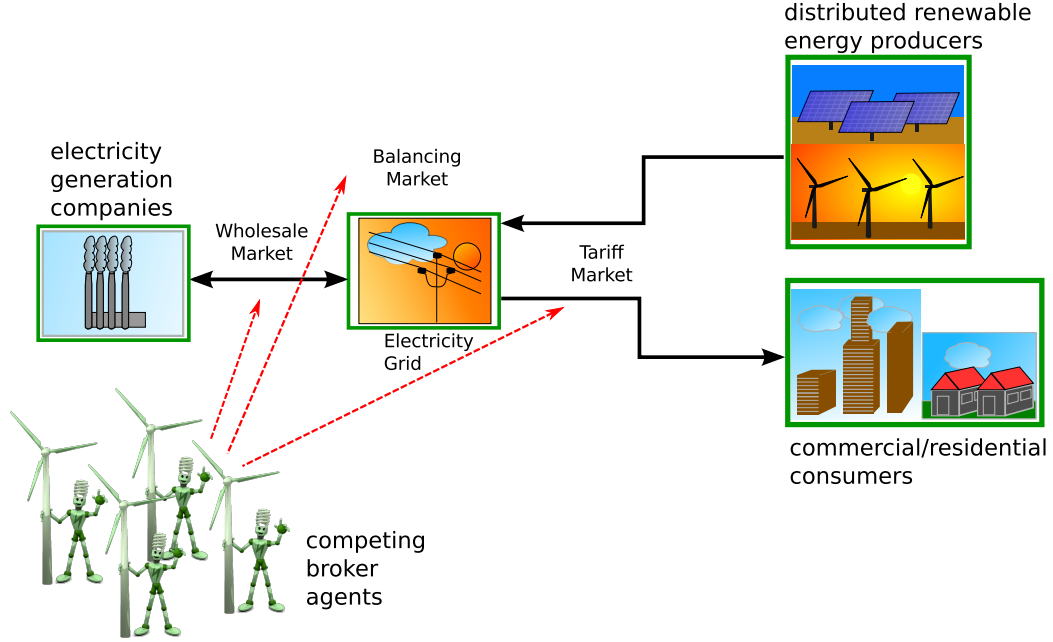


Figure 2.1: Structure of the Power TAC simulation environment

In Power TAC, autonomous broker agents compete by acting in three markets: (1) a *tariff market*, which is a retail market where energy is traded with consumers and distributed renewable energy producers, (2) a *wholesale market*, in which generation companies sell energy and brokers procure energy (or sell surplus), and (3) a *balancing market*, which ensures that electricity supply and demand are balanced at all times and determines broker imbalance fees. The brokers compete to maximize profit by gaining market share and trading electricity. A broker's game-score is its cash balance at the end of the game. A broker's total score in a competition is computed as a sum of its z-scores in each of the game-sizes, where a z-score in a game size is computed based on brokers' cumulative score in all games of this size. The simulation proceeds in 1-hour timeslots for about 60 simulated

days (about  $60 \times 24 = 1440$  timeslots). Each simulated timeslot takes 5 seconds of real-time, and the whole simulation takes about 2 hours to complete. Therefore, in each timeslot a broker has at most 5 seconds to complete its computation.

In the tariff market, brokers publish tariff contracts for energy consumption/production. Tariffs may include fixed and varying prices and possibly bonuses and/or fees. Once a tariff is published, customers can subscribe to it and consume/produce energy under this tariff as long as it is active (i.e. not revoked), paying or getting paid according to the tariff's terms. Customers stochastically subscribe to tariffs which maximize their utility, i.e. minimize their cost (or maximize their profit) and minimize their discomfort. A discomfort is created when a customer needs to shift its consumption to save costs (see an exact formulation in Section 7.1.2). Customers are equipped with smart-meters, so consumption and production are reported to the broker every hour. Some customers represent whole populations (e.g. a village of 30,000 people) and can subscribe subsets of their populations to different tariffs. Brokers may publish one or more tariffs once every 6 hours, 4 times per day.

In the wholesale market, brokers directly interact with each other, as well as with generation companies and other wholesale participants. The wholesale market is a day-ahead market that operates as a periodic double auction (PDA). It represents a traditional energy exchange, such as Nord Pool, or EEX.<sup>2</sup> At every hour, 24 independent double-auctions are executed in parallel, where each auction results in power to be delivered in one of the following 24 hours (timeslots). In these auctions, brokers, generation companies, and other simulated wholesale buyers place bids specifying amount to buy/sell, limit-price, and delivery time. Therefore, brokers have 24 opportunities to trade energy for each future timeslot, starting 24 hours in advance. Typically, brokers trade in the wholesale market to balance their

---

<sup>2</sup>See [www.nordpoolspot.gov](http://www.nordpoolspot.gov), [www.eex.com](http://www.eex.com)

portfolio’s net demand.

Power TAC’s wholesale market is a relatively simple call market that is modeled after existing wholesale power markets, with several simplifications. First, the effects of transmission constraints on auctions are not modeled: Power TAC’s wholesale market models a single region, with locational-marginal pricing modeled through a simple manipulation of the wholesale supply curve [46]. Second, while auctions for different timeslots are independent in Power TAC, in real markets they are coupled through unit-commitment and other issues. Third, in North-American and other real-world day ahead markets, the auctions do not have a rolling hourly structure such as the one in Power TAC; instead, bids for the next day are submitted once, on the previous day.<sup>3</sup> On the other hand, Power TAC’s wholesale market provides a realistic supply curve and auction clearing mechanism, which results in a realistic financial impact on brokers. For example, the wholesale market incentivizes brokers to reduce peak-demand and to contract with renewable producers. Moreover, since Power TAC does not model “real-time” markets (as discussed later in this section) the rolling structure of 24 sequential auctions provides an alternative that allows brokers to control their imbalance in close to real-time.

On the electricity grid, electricity supply and demand must be balanced at all times. The *balancing market* is responsible for real-time balancing of supply and demand. The balancing market provides a financial incentive for brokers to balance supply and demand in their portfolios in each timeslot, by being the most costly to use for balancing purposes. In the real world, this function is typically handled through ISO/TSO organizations and their ancillary services markets [9]. Since Power TAC does not model the full grid hierarchy, the balancing market provides

---

<sup>3</sup>Several European markets have intra-day markets that allow for some of the repeated adjustment of positions that can happen in Power TAC’s sequential auction mechanism. Still, these repeated adjustments are different from those in Power TAC. For example, the Spanish market has day-ahead bidding (i.e. 24 hourly, parallel auctions), and then three additional intra-day auctions; however, these intra-day auctions do not extend past the end of the day, so there is no continuous bidding process for the subsequent 24 hours such as the one in Power TAC.

a simplified version of the reserve and regulating capacity markets and associated controls normally operated by ISO/TSO [46].

Power TAC also models a Distribution Utility, which represents a regulated electric utility that operates the distribution grid. The Distribution Utility has two roles in Power TAC. First, it charges brokers a fixed distribution fee for the energy transported over the grid by their customers. Second, it provides “default tariffs” to which consumers and retail producers are subscribed in the beginning of a game. In this role, the distribution utility simulates a monopoly that exists prior to market liberalization. These default tariffs also bound brokers’ profits, since customers are always free to choose them over brokers’ tariffs. These default tariffs are analogous to the price-to-beat retail tariff that was required to be offered by incumbents in the ERCOT retail market when it first opened.

The state of the game is rich and high-dimensional, containing 100s to 10000s of variables (see Section 3.2.1): it includes the set of all active tariffs and customer subscriptions, the current energy consumption/production of all customers, the wholesale market deliveries and orders of all brokers for the following 24 hours, the internal states of all participants (brokers, customers, and generation companies), the current weather and weather forecast, the current time, and the bank balance of all brokers. The game state is partially observable to brokers. For instance, brokers sees all published tariffs in the tariff market but they only know the customer subscriptions for their own tariffs. Similarly, when an auction clears in the wholesale market, brokers only see the clearing price of the auction and a list of uncleared orders, but they do not know the list of cleared orders, or the future deliveries of other brokers. The action space of brokers is also high-dimensional, containing 10s to 100s of variables (see Section 3.2.1). For instance, tariff publications can include up to  $7 \times 24 = 168$  hourly energy prices, and wholesale market actions can include 24 sets of orders, one set for each of the following 24 hours.

## 2.2 Broker Interaction with the Environment

Autonomous agents operate by repeatedly sensing, deciding, and acting. This section describes in detail how brokers sense the environment and what actions are available to them. Later chapters of this dissertation focus on the broker’s decision making process. Figure 2.2 provides a simplified overview of the sequence of interactions of a broker with the simulation environment in each timeslot. The sequence of simulation processes described in the figure is fixed, while brokers can send messages at any time. Each simulation process uses all relevant broker messages that had arrived before it needs them. Other messages are used in the following timeslot. The following sections detail the available broker actions, the environment’s response to these actions, and the broker’s sensing of the environment.

### 2.2.1 Broker Actions

This section describes the actions available to brokers, focusing on actions used by brokers in this dissertation. In the tariff market, brokers can publish and revoke tariffs. In the wholesale market, brokers can submit limit orders to procure and sell energy for a future timeslot. In the balancing market, brokers can submit balancing orders to curtail consumption of interruptible customers, and adjust the consumption of storage customers for supply-demand balancing purposes. In our experience the current simulator with the current parameters does not emphasize balancing actions as being important strategically. Therefore the agents described in this dissertation do not consider them to any meaningful extent.

#### Tariff Market Actions

Table 2.1 lists the tariff market actions used in this dissertation. In the tariff market, brokers can publish one or more tariffs, revoke one or more tariffs, or take no action. A tariff is a contract for buying or selling energy. A tariff can generally be

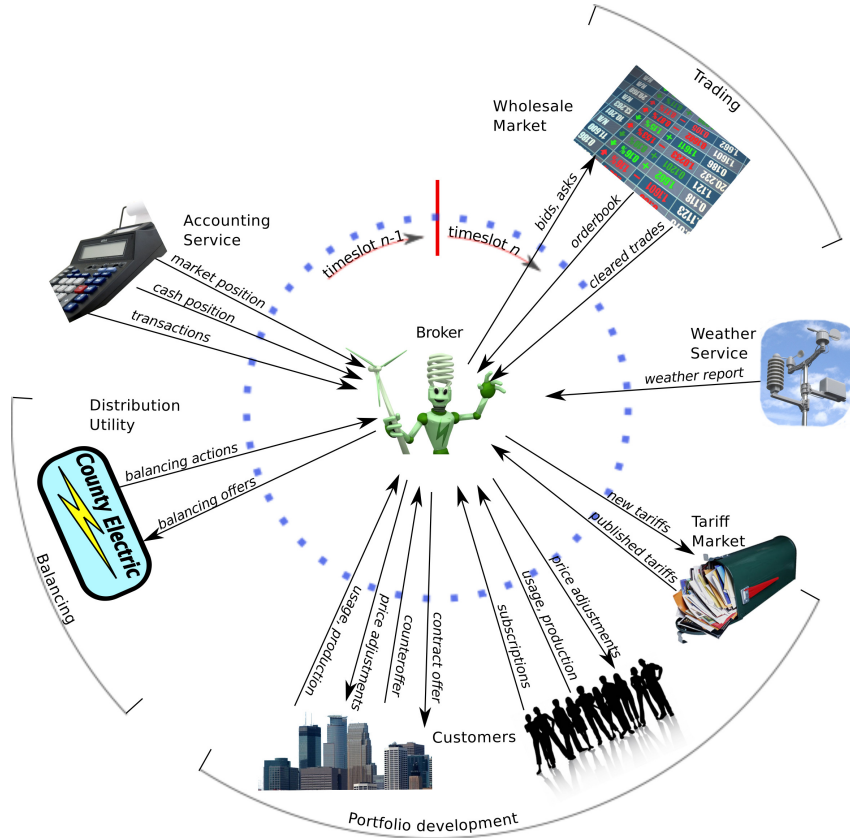


Figure 2.2: **Broker interactions with the simulation environment.** Source: the Power TAC game specification [46]. The diagram provides a simplified overview of Power TAC the interactions between a broker and the environment within one timeslot. Outgoing arrows represent possible actions, and incoming arrows represent information sent by the environment to the broker. In the wholesale market, a broker submits bids and asks for up to 24 parallel double auctions, and then receives the results of these auctions in the form of order-books and cleared orders. The weather service sends the current weather and forecast. In the tariff market, a broker can publish tariffs, receive a notification on tariffs published (by any broker), and receives updates on consumption/production of its customers as well as subscriptions to its tariffs. Negotiation with large customers is not implemented yet. In the balancing market, a broker can submit bids to balance its portfolio (e.g. through curtailment), and it receives information about the balancing actions executed and their costs. The accounting service sends the broker its wholesale market position, its cash position, and payment information for transactions executed in any of the markets.



a *consumption tariff* which is offered to consumers, or a *production tariff* which is offered to producers. Tariffs can also be more specific, e.g. a wind-production or solar-production tariff. As long as a tariff is active (i.e. published and not revoked), customers can subscribe to it, consume/produce energy and pay/get-paid according to its terms (See more details below, in Section 2.2.2). There is no limit on the number of tariffs a broker can publish and revoke at any given time, however publishing and revoking a tariff incurs non-negligible fees on the broker, which discourages the broker from flooding the market with tariffs. The tariff publishing and revoking fees are determined as game parameters.

Table 2.1: **Tariff market actions used in this dissertation.**

Action	Description
<code>publishTariff(<math>T</math>)</code>	publish tariff $T$ in the tariff market, making it available for all customers to subscribe
<code>revokeTariff(<math>T</math>)</code>	revoke tariff $T$ from the tariff market, making it unavailable for customers to subscribe
<code>no-op()</code>	Take no action.

Power TAC supports different types of tariffs, such as fixed-rate tariffs, Time-Of-Use tariffs, real-time price tariffs, tiered rate tariffs, and interruptible-rate tariffs. Each of these tariffs can be optionally augmented with signup, withdraw, and periodic payments. In this dissertation we use two of the above tariffs, which we have found to be most effective and most stably implemented and tested in the current implementation of the simulator: fixed-rate tariffs and Time-Of-Use tariffs, defined in Table 2.2. We do not use signup, withdraw or periodic payments.

## Wholesale Market Actions

Table 2.3 lists the available wholesale market actions. In the wholesale market, a broker can submit limit orders called bids and asks to buy and sell energy to be delivered in one of the following 24 hours, or take no action. For each of the following

Table 2.2: **Tariffs used in this dissertation.**

Action	Description
Fixed-rate tariff	A tariff $T := \langle type, rate \rangle$ , where $type \in \{\text{consumption, production, solar-production}\}$ determines what type of customers can subscribe to $T$ , and $rate$ specifies a fixed-price $p \in \mathbb{R}$ per-kWh of energy, applicable at any time.
Time-Of-Use (TOU) tariff	A tariff $T := \langle type, rate1, rate2, \dots \rangle$ , where $type$ is as for fixed-rate tariffs, and each of the (two or more) rates $rate1, rate2, \dots$ specifies a fixed-price $p \in \mathbb{R}$ per-kWh of energy that applies only during specific day/times. A valid Time-Of-Use tariff must cover each hour of the week with exactly one rate.

24 hours, a separate double-auction is executed with all orders submitted for this hour by brokers, generation companies, and other simulated wholesale buyers. The double-auction determines the cleared orders (energy quantities and prices). There is no limit on the number of orders a broker can submit for a future timeslot, other than the physical limit determined by the communication channel's bandwidth. However, orders with quantities smaller than 10 kWh are ignored, to prevent brokers from flooding the auction with infinitesimal orders. Submitting an order does not incur any fee on the broker.

### 2.2.2 Environment Response to Broker Actions

This section describes in more detail the simulation environment's (i.e. the endogenous agents') responses to brokers' actions, in the tariff market and in the wholesale markets.

Table 2.3: Wholesale market actions.

Action	Description
$\text{bid}(\langle e, l, t \rangle)$	A bid $\langle e, l, t \rangle := \langle \text{energyAmount}, \text{limitPrice}, \text{targetTime} \rangle$ specifies an energy amount $\text{energyAmount} \in \mathbb{R}_{>0}$ to procure, a limit-price $\text{limitPrice} \in \mathbb{R} \cup \{\infty\}$ that the broker is willing to pay per-MWh of energy ( $\infty$ is known as a <i>market-order</i> , which agrees for any price), and a target timeslot in one of the following 24 hours $\text{targetTime} \in \{+1, \dots, +24\}$ in which the energy is to be delivered. A convention of Power TAC is to precede a bid with a negative sign to signify that the bidder pays.
$\text{ask}(\langle e, l, t \rangle)$	An ask $\langle e, l, t \rangle := \langle \text{energyAmount}, \text{limitPrice}, \text{targetTime} \rangle$ specifies an energy amount $\text{energyAmount} \in \mathbb{R}_{<0}$ to sell, a limit-price $\text{limitPrice} \in \mathbb{R} \cup \{-\infty\}$ that the broker is willing to get paid per-MWh of energy (typically positive, meaning the broker is paid; $-\infty$ is known as a <i>market-order</i> , which agrees for any price, even if negative), and a target timeslot in one of the following 24 hours $\text{targetTime} \in \{+1, \dots, +24\}$ in which the energy is to be delivered.
$\text{no-op}()$	Take no action.

## Tariff Market Response to Broker Actions

In the tariff market, customers respond to brokers' tariff publications/revocations by (a) potentially subscribing to a new tariff, and (b) consuming or producing energy under the new tariff's terms.

**Subscribing to a Tariff.** Customers are approximate utility-optimizers: their subscription decisions are based on their predictions of their utilities under each candidate tariff. Specifically, tariff subscription is a two step process. First, customers predict their utility under each candidate tariff in the tariff market. Their utility is a weighted sum of monetary payments and discomfort factors. To predict their monetary payments under a tariff, they predict their expected consumption/production under this tariff, and compute the resulting payments based on the tariff terms. Payments may include one-time fees/bonuses e.g. for early-withdraw or signup, respectively. Discomfort factors take into account the need to shift consumption to save costs under non-fixed-rate tariffs (See a detailed description in Section 7.1.2), the need to switch brokers or tariffs, and the need to agree to consumption curtailment: all these factors reduce the desirability of a tariff by customers. Second, once customers compute the utility of each candidate tariff in the market, they use a probabilistic softmax selection rule between tariffs, where a tariff with higher utility is more likely to be selected. The softmax rule models imperfect tariff information in the real-world.

In Power TAC, about 90% of the consumption and more than 90% of the renewable production is done by customer models called *factored-customers* [87]. Some factored-customer models represent populations (e.g. a village of 30,000 people) and can subscribe subsets of their population to different tariffs. Other customer models represent an individual. In either case, each individual in the population is always subscribed to exactly one tariff at any given time. Customers *inertia* which

prevents customers from evaluating/switching tariffs too frequently.

**Consuming/Producing Under a Tariff.** Once customers are subscribed to a tariff, they consume/produce under this tariff. Customers' consumption/production is determined by some or all the following factors: weather conditions, time-of-day, day-of-week, elastic adaptation to tariff prices, random factors, shifted-consumption under non-fixed-rate tariffs. These factors have different impact on different customer types. For example, a residential customer will mainly consume in mornings, evenings and weekends, a commercial customer will mainly consume during weekdays, and a solar producer will produce in sunny days.

### **Wholesale Market Response to Broker Actions**

In the wholesale market, brokers' orders enter into double auctions, and auction results are sent back to brokers. At every hour, 24 independent double-auctions are executed in parallel, where each auction results in power to be delivered in one of the following 24 hours (timeslots). In these auctions, brokers, generation companies, and other wholesale buyers place bids/asks specifying a quantity to buy/sell, a limit-price, and delivery time. In the clearing process, demand and supply curves are constructed from bids and asks, and their intersection determines the clearing price and cleared quantity.

Figure 2.3 provides an example of a double auction. Note that bids specify a positive energy quantity and a negative price, and asks specify a negative energy quantity and a positive price. Note also that market orders (here specifying only quantity) are ordered first, since they agree at any price. In this example, there is no unique price where supply and demand curves intersect, therefore the average of these prices is taken to be the clearing price. All bids higher than the last executed bid and all asks lower than the last executed ask are fully executed for the same clearing price. In this auction-example the clearing price is 16, and the total cleared

quantity is 27 MWh. The last executed ask or bid may be only partially executed. In our example, the last ask is partially executed.

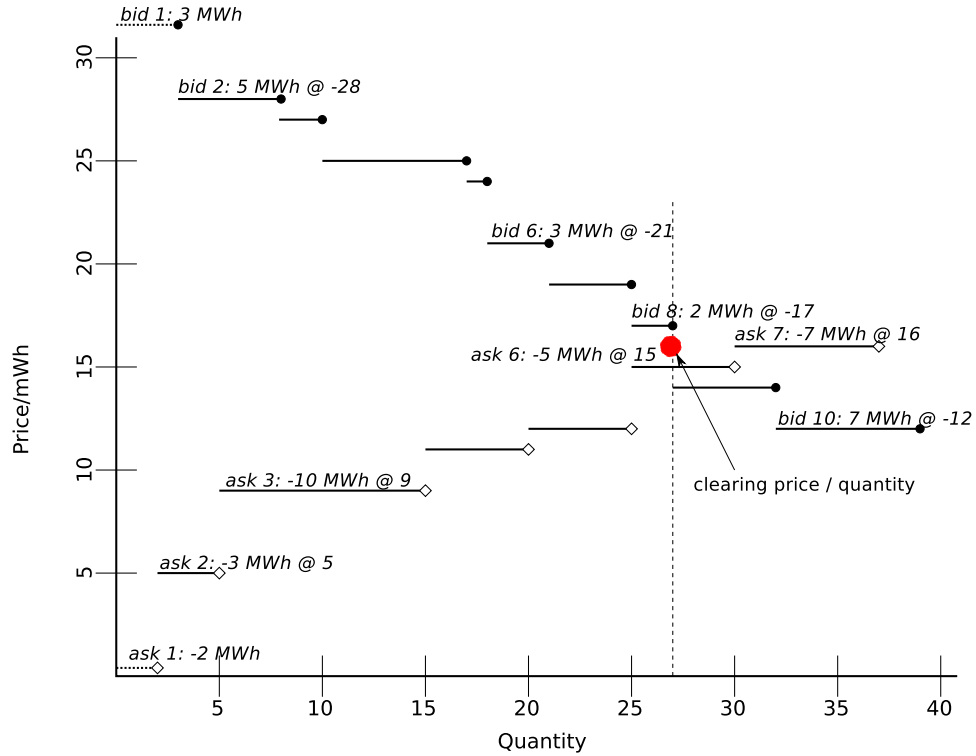


Figure 2.3: **Double-Auction Clearing Example.** Source: the Power TAC game specification [46]. See text for details.

After the auction completes, the following information is sent to brokers:

- The clearing price (16) and total quantity (27 MWh) are sent to all brokers.
- The uncleared orders are sent to all brokers, without broker information. In this case these are the asks (-3, 15) and (-7, 16) and the bids (5, -14) and (7, -12).
- Each broker's cleared orders are privately sent to the broker.
- Updated cash and wholesale market positions (future energy deliveries for the timeslot for which the auction was executed) are privately sent to each broker.

- All orders that participated in the auction are discarded.

### 2.2.3 Broker Sensing

In Power TAC, a broker senses the environment through messages sent to it by the simulation environment. These messages reveal to the broker partial information about the simulation’s world state. This section lists the types of messages sent to a broker by the simulator. We group messages by whether they are public (sent to and known by all brokers) or private (sent to and known by a single broker), and by the times in which they are sent. This section closely follows Section 3.3 from the Power TAC 2015 game description [46].

#### Public Messages Sent by the Power TAC Simulator

The following messages are sent publicly to all brokers *at the beginning of a game*:

- **Game Parameters:** the parameters used to configure the current game. The Power TAC game parameters are listed in Appendix B.
- **Competing Broker Identities:** the identities of competitor brokers, which remain unchanged throughout a competition.
- **Customers List:** names and properties of customers in the current game, most importantly their type (consumers, producers, solar-producers, interruptible consumers, etc.) and population-size.
- **Default Tariffs:** tariffs offered by the distribution utility, to which customers are subscribed at the beginning of the game, and can re-subscribe at any point in the game. Typically there are two such tariffs, one for producers and one for consumers, at default prices that would be used in a monopoly situation.
- **Bootstrap Customer Data:** consumption and production of each customer in the 14 days prior to the beginning of the game, under the default tariffs.

- **Bootstrap Wholesale Market Data:** delivered prices and energy quantities by generation companies to the customers of a default broker that represents a monopoly in the 14 days prior to the beginning of the game.
- **Bootstrap Weather Data:** weather conditions in the 14 days prior to the beginning of the game.

The following messages are sent publicly to all brokers *once every 6 simulated hours*:

- **Tariff Updates:** updates on newly published, revoked, or modified tariffs by all brokers.

The following messages are sent publicly to all brokers *once every simulated hour*:

- **Wholesale Market Clearing Data:** wholesale market total quantities traded and clearing prices for each of the 24 auctions executed in this timeslot.
- **Wholesale Market Order Books:** lists of uncleared bids and asks (each includes a price and a quantity) from each of the 24 auctions executed in this timeslot.
- **Total Consumption and Production:** the total consumption and the total production by all simulated customers in the current timeslot.
- **Weather Report and Forecast:** weather conditions for the current timeslot (temperature, cloud cover, wind direction, wind speed), and a 24-hour forecast of these conditions.

#### **Private Messages Sent by the Power TAC Simulator**

The following messages are sent privately to each broker *once every 6 simulated hours*:



- **Tariff and Subscription Changes Transactions:** confirmations and paid fees for the brokers' tariff publication and revocations, subscription changes of the brokers' customers (either sign-up or withdraw) and associated payments if exist, such as signup and early withdraw payments.

The following messages are sent privately to each broker *once every simulated hour*:

- **Tariff Consumption/Production and Payment Transactions:** consumption and production of the brokers' customers in the current timeslot and the associated payments, broken down by customer-subscriptions (customer-tariff pairs).
- **Wholesale Market Transactions:** cleared and partially cleared bids and asks submitted by the broker.
- **Wholesale Market Positions:** energy to be delivered to/by the broker in each of the following 24 timeslots.
- **Distribution Transactions:** the quantity distributed by the broker among its customers and the associated charges for this distribution.
- **Balancing Transactions:** the broker's supply-demand imbalance and the associated charges for this imbalance.
- **Cash Position:** the broker's current bank balance.

## 2.3 Power TAC Modeling Assumptions

This section lists the modeling assumptions used by the Power TAC simulation environment [46], as follows.

1. Line capacity limitations are ignored, reflecting an assumption that electricity flows unconstrained over the simulated distribution grid, among all participants. This assumption frequently holds in distribution grids, so it is not a major restriction. Once more distributed generators and storage facilities will be simulated, this assumption will need to be re-examined.
2. Power factor effects, i.e. phase shifts between voltage and current, are ignored. A lower power factor typically results in higher energy losses over the distribution network. Therefore, electric utilities typically charge a higher cost to industrial or commercial customers with a low power factor. Modeling power factor effects in Power TAC may affect brokers' decisions on how to charge customers, but is currently out of scope.
3. Electricity distribution and transformation losses are ignored. These losses are estimated to be 5.5% in North America [115], and can be considered roughly constant and similar among distribution grid participants. Therefore, the validity of the simulation results is not affected by this assumption.
4. In addition to traditional generation companies, two types of producers (electricity production facilities) are simulated. Producers of the first type produce electricity whenever they are active. For example, solar arrays and wind turbines are activated by weather conditions. Producers of the second type are "controllable": their output can be adjusted remotely within their capacity range. Examples of this type are electric vehicle batteries, and some combined heat and power units. Both of these producer types are becoming increasingly common in the real world, and are therefore included in the simulation.
5. Real-time operations of the local distribution grid, including supply-demand balancing, are outside the control of competing brokers, and are executed using a combination of controllable generators and spinning reserves. In the

real-world these operations are typically managed by a system operator.

6. Brokers pay for supply-demand balancing as determined by a *balancing-market*, which constitutes a simplified version of real-world reserve and regulating capacity markets. Power TAC is not intended to model these markets in detail; the balancing-market’s goal is mainly to incentivize brokers to balance supply and demand as closely as possible.
7. Simulation time progresses in discrete timeslots, each representing one hour. Timeslots correspond to trading intervals in a regional wholesale market. This assumption allows for simulating a period of months rather than minutes or hours. However, under this assumption, the temporal distribution of energy consumption and production cannot be captured. For example, supply-demand imbalance is computed as the difference between total consumption and production in a one-hour timeslot, rather than as an instantaneous difference between two continuous time-series.<sup>4</sup>
8. Some portions of consumption can be controlled directly (through curtailment) or indirectly (using price signals). In the latter case, autonomous agents adjust consumption on behalf of human consumers to optimize cost and comfort. This assumption currently holds to some extent in the real-world, and is expected to hold more widely in the future. For example, programs that allow for consumption curtailment by utilities exist (e.g. in Austin, Texas), and some autonomous agents are already being installed in buildings (such as smart-thermostats).

---

<sup>4</sup>In the US and in some European markets, in addition to reserves and regulating markets there are also “real-time” markets that allow for another set of auctions, at 5 or 15 minute intervals (where offers are updated every hour and stay fixed for all 12 of the 5-minute auctions in any given hour). The real-time market deals with some of the intra-hour supply demand balance, with the result that less of the balance needs to be dealt with by the reserves and regulating capacity markets. Such “real-time” markets are not modeled in Power TAC, and are assumed to be included in the balancing-market model. However, the rolling structure of Power TAC’s 24 sequential double auctions provides an alternative that allows brokers to control their imbalance in close to real-time.

To summarize, the Power TAC simulator is fairly detailed and realistic, and therefore we contend that it can be viewed as a reasonable substrate domain for studying general electricity market conditions.

## 2.4 Chapter Summary

Future smart grids will require new power market structures [114]. Due to the risk of deploying new market structures in the real-world [7], it is important to test new market structures first in simulation. The Power TAC simulator is a detailed and realistic retail markets simulator, and therefore we contend that it can be viewed as a reasonable substrate domain for studying general electricity market conditions. We use the Power TAC simulation environment as a substrate domain throughout this dissertation.

In Power TAC, autonomous brokers compete with other for making profit by gaining market share and trading energy with about 57,000 retail customers (consumers and producers), by trading energy with generation companies and other brokers in a wholesale market, and by participating in a real-time supply-demand balancing-market. Such autonomous, self-interested, broker agents can be financially incentivized to contribute to social welfare, for example by maintaining supply-demand balance in their portfolios, thus contributing to grid stability. We overviewed the Power TAC simulator and detailed the parts that are used in the rest of this dissertation. Full details can be found in the Power TAC Game Specification [46]. In the rest of this dissertation, we focus on the decision making process of an autonomous broker agent in smart grid environments, as modeled by Power TAC.

# Chapter 3

## The Broker's Power Trading Problem: Formalization

This chapter formalizes the complete broker's power trading problem. Our formalization is beneficial in multiple ways, as it: (1) compactly captures the complex challenges faced by a broker, (2) characterizes the optimal solution, (3) provides a guideline for approximating the solution and for extending existing approximate solutions. Indeed, in the following chapters we use our problem formalization to design a series of approximate solutions of increasing generality, which work effectively in practice. While our formalization is based on the Power TAC simulator, we expect it to generalize and be useful in reality, since Power TAC closely models real-world markets. We start with an intuitive problem description (Section 3.1), then formalize the power trading problem as a Markov Decision Process (Section 3.2), and characterize approximate solutions (Section 3.3).

### 3.1 The Power Trading Problem’s Temporal Structure

Figure 3.1 illustrates the temporal structure of a broker’s power trading problem. In our illustration and in the rest of the dissertation we exclude balancing market actions, since we have found in preliminary tests that the current simulator version does not provide enough incentive to use them. Henceforth, we focus on tariff and wholesale market actions. The temporal structure of the tariff and wholesale market actions differ in multiple ways. Tariffs specify energy for *immediate* and *repeated* delivery and are published at *low-frequency* (every one or more days). Immediate delivery means that once a tariff is published, customers can subscribe to it immediately and consume/produce energy under its payment terms. Repeated delivery means that customers can do so indefinitely, until the tariff is either revoked or expires. Tariff publications/revocations are typically executed at low frequency (every one or more days) since (1) tariffs are for repeated delivery, (2) each tariff publication/revocation incurs a fee paid by the broker, and (3) customers respond slowly to tariff publications, due to inertia that binds them to their current tariff.

In contrast to the tariff market, wholesale actions (bids/asks, referred to here as “orders”) typically specify energy for *future, one-time* delivery and are executed at *high-frequency* (every hour). Future delivery means that the energy and payment transactions specified in a cleared order happen in the future (one of the next 24 hours). One-time delivery means that these transactions happen once, in a single timeslot. Since wholesale orders are for one-time delivery, and since the broker trades energy continually, wholesale orders are executed at high-frequency.

### 3.2 Power Trading as a Markov Decision Process

Given the internal states of the simulator and competing brokers, the broker’s energy trading problem is a Markov Decision Process (MDP) [82]. However, since

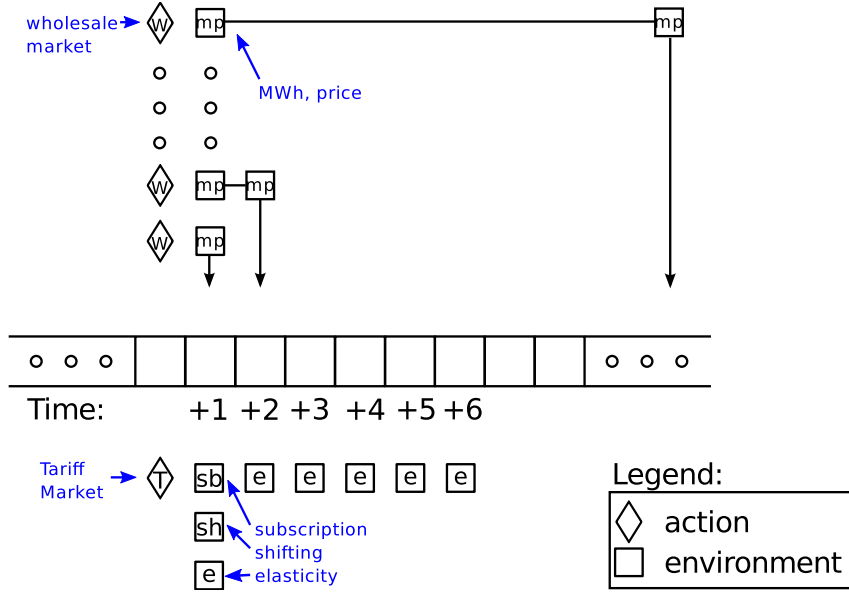


Figure 3.1: **Temporal structure of the power trading problem.** Time progresses to the right; the notation ‘ $+i$ ’ stands for ‘ $i$  timeslots into the future’. Diamonds stand for broker actions. Squares stand for simulation environment responses. The top part represents the wholesale market: a broker submits limit orders to buy/sell energy for the next 24 hours, then it receives the results of the 24 double-auctions. The bottom part represents the tariff market: a broker may publish one or more tariffs (once every 6 hours), and customers respond by potentially (1) subscribing to new tariffs, (2) shifting consumption to cheaper times, and (3) elastically adapting total consumption based on price.

competitors’ state and parts of the simulator state are unobservable, the trading problem is actually a Partially Observable MDP (POMDP). Nevertheless, for computational tractability and modeling clarity, we approximate the trading problem as an MDP. We present our MDP definition next, denoting the acting broker as  $B_0$ . For easier reference, Table 3.1 summarizes the variables used for defining the MDP’s main components, Table 3.2 summarizes the variables used in the MDP’s state definition, and Table 3.3 summarizes the main variables used in the MDP’s reward definition.

Table 3.1: **Variables used for the power trading MDP’s main components.**

Variable	Description
$S$	the set of states
$A$	the set of all actions
$A^\tau$	the set of tariff market actions
$A^\omega$	the set of wholesale market actions
$A^\beta$	the set of balancing market actions
$\gamma$	discount factor
$s_t$	state at time $t$
$a_t$	action(s) taken at time $t$
$r_t$	reward at time $t$
$r^\tau$	tariff market component of the reward
$r^\omega$	wholesale market component of the reward
$r^\beta$	balancing market component of the reward

- **States:**  $S$  is the set of states, where state  $s$  is a tuple

$$\langle t, \mathcal{B}, \mathcal{C}, \mathcal{P}, \mathcal{T}, \mathcal{S}_{B_0}, \mathcal{Q}_{B_0}, \mathcal{A}_{B_0}, I_{B_0}, \mathcal{W}, \$_{B_0}, \mathcal{R} \rangle$$

that includes the current time  $t$  (which encapsulates weekday/hour), and the sets: competing broker identities  $\mathcal{B}$ ; identities of consumers  $\mathcal{C}$  and producers  $\mathcal{P}$  (both referred to as *customers*); published tariffs of all brokers  $\mathcal{T} := \cup_{B \in \mathcal{B}} \mathcal{T}_B$ ; customer subscriptions to  $B_0$ ’s tariffs  $\mathcal{S}_{B_0}$ ; current energy consumption/production of  $B_0$ ’s customers  $\mathcal{Q}_{B_0}$ ; recent auction results  $\mathcal{A}_{B_0} := \{\langle p^c, q^c, \mathcal{O}^c_{B_0}, \mathcal{O}^u, \mathcal{M}_{B_0} \rangle_j\}_{j=t+1}^{t+24}$  including, for each of the following 24 timeslots, the clearing price  $p^c$  and total quantity  $q^c$ ,  $B_0$ ’s cleared orders  $\mathcal{O}^c_{B_0}$ , all brokers’ uncleared orders  $\mathcal{O}^u$ , and  $B_0$ ’s wholesale *market-positions*  $\mathcal{M}_{B_0}$  (energy deliveries and charges, updated incrementally from  $\mathcal{O}^c_{B_0}$ );  $B_0$ ’s energy imbalance  $I_{B_0}$ ; current weather and forecast  $\mathcal{W}$ ;  $B_0$ ’s cash balance  $\$_{B_0}$ ; and randomly sampled game-parameters (such as fees and game length)  $\mathcal{R}$ .

**Note:** the underlying state of the game, which includes elements unobserved



Table 3.2: **Variables used in the power trading MDP’s state definition.** We use the following conventions: calligraphic letters (e.g.  $\mathcal{B}$ ) stand for sets; uppercase letters stand for items in sets (e.g.  $B$  is an item in  $\mathcal{B}$ ); lowercase letters are items or numbers that are not part of a set; a superscript symbolically augments a variable’s letter (to avoid two-letter variables, for instance,  $p^c$  is an auction’s clearing price, where the superscript clarifies the type of the price); a subscript defines either a subset (when it indexes a calligraphic letter), or an item in a set (when it indexes an uppercase letter). One exception is our use of  $\$$  to stand for a set of cash balances, and of  $\$B$  to stand for one item in the set, due to the symbolic meaning of the sign  $\$$ .

**Note:** these conventions could lead to conflicts with other notation in this dissertation (e.g. MDP states and a subscription share the letter  $S$ ). Due to the large number of variables it is hard to avoid some conflicts, so in case that happens, we make sure to explicitly disambiguate and clarify the use of a variable. We preferred the option of keeping notation intuitive at the expense of some notation overload, rather than using unique, but non-intuitive letters for variables.

Variable	Description
$t$	current time (which encapsulates weekday/hour)
$\mathcal{B}$	the set of competing brokers’ identities
$\mathcal{B}^o$	the set of competing brokers, including their internal states
$B$	a broker identity
$B_0$	the identity of the broker acting in the MDP
$\mathcal{G}^o$	the set of generation companies, including their internal states
$\mathcal{C}$	the set of consumers’ identities
$\mathcal{C}^o$	the set of consumers, including their internal states
$\mathcal{P}$	the set of producers’ identities
$\mathcal{P}^o$	the set of producers, including their internal states
$\mathcal{T}$	tariffs published by all brokers
$\mathcal{T}_B$	tariffs published by broker $B$
$\mathcal{S}$	customer subscriptions
$\mathcal{S}_B$	customer subscriptions to broker $B$ ’s tariffs
$\mathcal{Q}$	current energy consumption/production of all brokers
$\mathcal{Q}_B$	current energy consumption/production of broker $B$ ’s customers
$\mathcal{A}$	recent auction results for all brokers
$\mathcal{A}_B$	recent auction results for broker $B$
$p^c$	auction clearing price
$q^c$	auction’s total traded quantity
$\mathcal{O}^c_B$	auction’s cleared orders of broker $B$
$\mathcal{O}^u_B$	auction’s uncleared orders of all brokers
$\mathcal{M}_B$	wholesale market-position of broker $B$
$\mathcal{I}$	energy imbalances of all brokers
$I_B$	energy imbalance of broker $B$
$\mathcal{W}$	current weather and forecast
$\$$	cash balance of all brokers
$\$B$	cash balance of broker $B$
$\mathcal{R}$	randomly sampled game-parameters (such as fees and game length)

Table 3.3: **Variables used in the power trading MDP’s reward definition.**

Variable	Description
$Q_t^{cons}$	energy quantity consumed by the broker’s customers at time $t$
$p_t^{cons}$	average price of energy sold to the broker’s customers
$Q_t^{prod}$	energy quantity produced by the broker’s customers at time $t$
$p_t^{prod}$	average price of energy procured from the broker’s customers
$Q_t^{ask}$	energy quantity sold by the broker in the wholesale market at time $t$
$p_t^{ask}$	average price of energy sold by the broker in the wholesale market
$Q_t^{bid}$	energy quantity procured by the broker in the wholesale market at time $t$
$p_t^{bid}$	average price of energy procured by the broker in the wholesale market

by the broker, is the tuple

$$\langle t, \mathcal{B}^o, \mathcal{G}^o, \mathcal{C}^o, \mathcal{P}^o, \mathcal{T}, \mathcal{S}, \mathcal{Q}, \mathcal{A}, \mathcal{I}, \mathcal{W}, \$, \mathcal{R} \rangle$$

where  $\mathcal{B}^o, \mathcal{G}^o, \mathcal{C}^o, \mathcal{P}^o$  are the sets of brokers, generation companies, consumers and producers, respectively (the difference from  $\mathcal{B}, \mathcal{G}, \mathcal{C}, \mathcal{P}$  is that while these sets included only object identities, the sets  $\mathcal{B}^o, \mathcal{G}^o, \mathcal{C}^o, \mathcal{P}^o$  include complete objects, and importantly the internal states of these objects)<sup>1</sup>; and where  $\mathcal{S} := \cup_{B \in \mathcal{B}} \mathcal{S}_B$ ,  $\mathcal{Q} := \cup_{B \in \mathcal{B}} \mathcal{Q}_B$ ,  $\mathcal{A} := \cup_{B \in \mathcal{B}} \mathcal{A}_B$ ,  $\mathcal{I} := \{I_B\}_{B \in \mathcal{B}}$ ,  $\$ := \{\$B\}_{B \in \mathcal{B}}$ .

- **Actions:** A broker’s set of actions  $A := A^\tau \cup A^\omega \cup A^\beta$  is composed of tariff market actions  $A^\tau$ , wholesale market actions  $A^\omega$  and balancing market actions  $A^\beta$ , as follows.

1. *Tariff market actions*  $A^\tau$ : publish/modify/revoke tariffs. A tariff is a tuple  $T = \langle type, rates, fees \rangle$  where:

- $type \in \{\text{consumption, production, ...}\}$  can be general (e.g. production)

---

<sup>1</sup>An internal state generally refers to the computation state of an object. Different objects may have different representations of their internal state. For example, two competing brokers may be implemented differently, and therefore their computation states may have different representations. Most generally, an object’s internal state is a snapshot of its representation in the computer’s memory. The distinction between an object identity and a complete object that includes an internal state is important for understanding our MDP: while the latter is part of the complete MDP’s state, only the former is observed by the broker acting in the MDP.

or specific (e.g. solar-production).

- *rates*: a set of rates, each specifying price/kWh and times, and/or usage thresholds where it applies.
- *fees*: optional periodic/signup/withdraw payments.

2. *Wholesale market actions*  $A^\omega$ : submit limit orders of the form

$$\langle \text{energyAmount}, \text{limitPrice}, \text{targetTime} \rangle$$

to buy/sell energy for one of the next 24 hours.

3. *Balancing market actions*  $A^\beta$ : submit customers energy curtailment requests (currently unused).

- **Transition Function:** The transition function is partially deterministic and partially stochastic, as follows. The time  $t$  is incremented by 1 hour;  $\mathcal{B}, \mathcal{C}, \mathcal{P}$  remain unchanged;  $\mathcal{T}$  is updated by publish/modify/revoke tariff actions, deterministically by  $B_0$ , and stochastically (due to unobservability) by other brokers;  $\mathcal{S}_{B_0}$  is updated stochastically based on customers' decisions;  $\mathcal{Q}_{B_0}$  is determined stochastically based on weather and customers' internal states (shifting and elasticity, see Figure 3.1);  $\mathcal{A}_{B_0}$  is updated with auction results, stochastically since (i) competitors rely on stochastic information (demand predictions), (ii) competitors' internal states are hidden, and (iii) generation companies bid stochastically;  $I_{B_0}$  is a deterministic function of  $\mathcal{T}_{B_0}, \mathcal{S}_{B_0}, \mathcal{Q}_{B_0}, \mathcal{A}_{B_0}$ ;  $\mathcal{W}$  is stochastic;  $\$$  is updated deterministically from the recent stochastic reward; and  $\mathcal{R}$  remains unchanged.
- **Reward:** Let  $s_t, r_t, a_t$  be the state, reward, and broker-action(s) at time  $t$ . Let  $r^\tau, r^\omega, r^\beta$  be the broker's energy buy/sell payments in the tariff, wholesale, and balancing markets respectively. Let *dist* be the energy distribution fees, and *fees* the tariff-market fees. The reward at time  $t$  can be characterized by

the following function.

$$\begin{aligned}
r_t(s_{t-1}, a_{t-1}, s_t) := & r^\tau(s_t) + r^\omega(s_t) + r^\beta(s_t) + dist(s_t) + fees(s_{t-1}, a_{t-1}, s_t) := \\
& \underbrace{Q_t^{cons} p_t^{cons} - Q_t^{prod} p_t^{prod}}_{r^\tau(s_t)} + \underbrace{Q_t^{ask} p_t^{ask} - Q_t^{bid} p_t^{bid}}_{r^\omega(s_t)} \underbrace{\pm bal(I_{B_0,t})}_{r^\beta(s_t)} \\
& \underbrace{-max(Q_t^{cons}, Q_t^{prod}) \times distFee}_{dist(s_t)} \underbrace{-pub(a_{t-1}) - rev(a_{t-1}) \pm psu(\mathcal{S}_{B_0,t-1}, \mathcal{S}_{B_0,t})}_{fees(s_{t-1}, a_{t-1}, s_t)}
\end{aligned} \tag{3.1}$$

where  $\pm$  denotes components that can be positive or negative;  $Q_t^{cons}, Q_t^{prod}$  are the total consumed/produced quantities by  $B_0$ 's customers in the tariff-market (both are sums of entries of  $\mathcal{Q}_{B_0}$ );  $Q_t^{ask}, Q_t^{bid}$  are the amounts  $B_0$  sold/procured in the wholesale-market (both are sums of elements of  $\mathcal{M}_{B_0}$  inside  $\mathcal{A}_{B_0}$ );  $p_t^{cons}, p_t^{prod}, p_t^{ask}, p_t^{bid}$  are the average buying/selling prices (determined by  $\mathcal{T}_{B_0}, \mathcal{S}_{B_0}, \mathcal{Q}_{B_0}$  and  $\mathcal{M}_{B_0}$ );  $bal(I_{B_0,t})$  is the fee for imbalance  $I_{B_0,t} = Q_t^{cons} - Q_t^{prod} + Q_t^{ask} - Q_t^{bid}$  (which depends on unobserved other broker imbalances  $I \setminus I_{B_0,t}$ );  $distFee$  is a fixed fee per kWh transferred over the grid;  $pub, rev$  are tariff publication and revoke fees;  $psu$  are tariff periodic/signup/withdraw fees/bonuses.

- **Discount Factor:**  $\gamma$  reflects daily interest on cash balance.

### 3.2.1 Dimensionality of the State and Action Spaces

To get a better understanding of the size of the trading problem, we estimate the number of dimensions (i.e. variables) in the MDP's state and action spaces. Frequently, complex MDPs use *factored state representations*, in which a state is represented as a fixed set of variables, each taking values from some discrete or continuous domain. However, in the power trading problem's MDP, a state cannot be represented as a fixed set of variables since it contains information of unbounded, variable

size (bounded only by the computer memory's size). Example of such variable-size information is the set of existing tariffs and the set of uncleared wholesale orders. In practice, the sizes of these sets are within some reasonable ranges, based on which we provide a rough estimation of the typical number of dimensions in a state, as follows.

- **Time** ( $t$ ): 1 variable.
- **Identities of brokers, consumers and producers** ( $\mathcal{B}, \mathcal{C}, \mathcal{P}$ ): typically 10s of variables.
- **Published tariffs** ( $\mathcal{T}$ ): typically 10s to 100s of tariffs, each containing 1-168 prices, 0-4 fee parameters, so that the number of variables is roughly 10s to 10000s.
- **Broker subscriptions** ( $\mathcal{S}_{B_0}$ ): between 0 and the number of customers. The number of customers is currently about 57,000, but in the current implementation customers are grouped into up to several 100s of groups (each member of a group consumes/produces identically), therefore the number of subscriptions is between 0 and 100s.
- **Current energy consumption/production of customers** ( $\mathcal{Q}_{B_0}$ ): between 0 and the number of customers which is, as was just described, up to 57,000 but currently typically 100s of groups.
- **Auction results** ( $\mathcal{A}_{B_0}$ ): 24 sets, each containing 0 to 10s of cleared and uncleared orders, and one market-position. Each order has 2 variables, so the total number of variables is around 100s.
- **Broker imbalance** ( $I_{B_0}$ ): 1 variable.

- **Weather and forecast** ( $\mathcal{W}$ ): 25 sets, each with 4 variables, describing the current weather conditions and the conditions in the following 24 hours. Therefore, the number of variables is 100.
- **Cash balance** ( $\$_{B_0}$ ): 1 variable.
- **Game parameters** ( $\mathcal{R}$ ): 24 variables (See Appendix B)

Based on these estimations, the number of state variables ranges between 100s to 10000s of variables. The action space is high-dimensional as well. A tariff action contains 1-168 prices and 4 fee parameters, and a broker can publish any number of tariffs at a given time. Historically, brokers have typically published a single tariff at a time, so the number of tariff-action dimensions is practically 1-172. A wholesale bid has two variables, and a broker can submit any number of bids. Historically, brokers have typically published 0-10s of bids at each time, so the number of variables of wholesale actions is around 10s.

### 3.3 Lookahead Policies as Approximate Solutions

The MDP's solution is an optimal power-trading policy (a mapping from states to actions). There are two problems that prevent solving the MDP exactly: first, the high-dimensional states and actions and the complex reward makes it computationally intractable, and second, some components of the transition and reward functions are unknown to the broker. Therefore, brokers necessarily can only approximate the solution.

Powell et al ([79]) identify four basic classes of approximate solutions to large MDPs: policy function approximation, cost function approximation, value function approximation, and lookahead policies. The effectiveness of each of these classes varies based on specific problem properties. Furthermore, in practice it might not be possible to test all of these policy classes for every problem, since some classes

might not be applicable to some problems. Therefore [79] proposes general guidelines for when to use each class, based on their experience:

- Policy function approximations are applicable when there is a clear relationship between state and action, or when the policy function can be approximated accurately.
- Cost function approximations are applicable when a deterministic model provides a good solution, where some adjustments are needed to make the solution more robust.
- Value function approximations are applicable when the value of being in a state is easy to approximate.
- Lookahead policies are needed in time-varying (i.e. non-stationary) settings, and when the value of being in a state is hard to approximate.

Based on these guidelines *lookahead policies* seem suitable for our domain, which is non-stationary due to factors like weather and other brokers’ policies, and where it is unclear how to approximate a value function accurately due to the high-dimensional state and action spaces and the complex reward function.

Lookahead policies are a class of partial solutions for MDPs and POMDPs (Partially Observable MDPs) that can be effective in high-dimensional state-spaces [6, 39, 22, 20, 64, 97, 109]. Lookahead policies come in different variations, and include *brute-force tree-search* policies, *Monte-Carlo tree search* policies, *roll-out heuristics*, *deterministic rolling-horizon procedures* (also known as *receding horizon procedures*, and *model-predictive control*), and *stochastic rolling-horizon procedures* [80].

Lookahead policies make a decision in a given state by solving an approximation of the problem over some horizon. Instead of computing the optimal policy over the entire state-space, they approximate it by optimizing over simulated hypothetical lookahead trajectories  $s_t, r_t, a_t, s_{t+1}, r_{t+1}, a_{t+1}, \dots$  using generative models that

predict *action effects* (next state and reward). In the trading problem’s MDP, the reward is a deterministic function of  $s_{t-1}$ ,  $a_{t-1}$ ,  $s_t$  except for the  $bal(I_{B_0,t})$  component. Therefore a broker needs generative models for  $bal(I_{B_0,t})$ , for  $\mathcal{T} \setminus \mathcal{T}_{B_0}, \mathcal{S}_{B_0}, \mathcal{Q}_{B_0}$  (to predict  $Q_t^{cons}, p_t^{cons}, Q_t^{prod}, p_t^{prod}$ ), and for  $\mathcal{A}_{B_0}$  (to predict  $Q_t^{ask}, p_t^{ask}, Q_t^{bid}, p_t^{bid}$ ).

While lookahead policies reduce the complexity of finding approximate MDP solutions in high-dimensional *state-spaces*, they can still be intractable in presence of high-dimensional *action-spaces*, if they blindly search over (subsets of) high-dimensional actions at every lookahead step. In our domain, while action effects can be predicted independently for each action type (specifically  $Q_t^{cons} p_t^{cons}, Q_t^{prod} p_t^{prod}, Q_t^{ask} p_t^{ask}, Q_t^{bid} p_t^{bid}$  for: consumption tariffs, production tariffs, wholesale sell, wholesale buy, respectively), (i) there is an innumerable set of possible (subsets of) high-dimensional actions of each type, and (ii) different action-types cannot be optimized independently: the  $bal(I_{B_0,t})$  function is designed such that imbalance fees typically result in negative reward when taking actions of a single type, while positive reward can be achieved by taking actions of multiple types in parallel (to maintain low imbalance). Therefore, any tractable lookahead policy is required to efficiently (i) sample, and (ii) combine the actions to simulate.

### 3.4 Chapter Summary

In this chapter, we formalized the broker’s power trading problem as an MDP. Due to the high-dimensional state and action spaces, it is computationally intractable to solve this MDP exactly. Of the different classes of approximate MDP solutions, lookahead policies seem appropriate for the power trading domain. In the next chapter, we describe one of the main contributions of this dissertation: the LATTE algorithm, which defines a general lookahead policy that approximates the MDP’s solution. Later chapters (specifically Chapters 5, 6, and 7) will instantiate this general algorithm in specific implementations that work effectively in practice.



## Chapter 4

# The **LATTE** Algorithm: Lookahead-policy for Autonomous Time-constrained Trading of Electricity

In the previous chapter, we asserted that lookahead policies have the potential to perform effectively in the power trading domain, and we pointed out two challenges that a tractable lookahead policy must address: efficiently (i) sampling, and (ii) combining the actions to simulate in a lookahead trajectory. This chapter introduces a general lookahead policy that approximates the power trading MDP’s solution, and which is used throughout this dissertation by all our agents: the LATTE algorithm (Lookahead-policy for Autonomous Time-constrained Trading of Electricity). LATTE addresses (i) and (ii) by utilizing functional dependencies in the broker’s action effects, and by utilizing the temporal structure of the power trading problem (Figure 3.1). In that sense, LATTE is specific to the power trading problem. On

the other hand, LATTE serves as a general template for power trading lookahead policies, by leaving parts of its definition open for different implementations. In that sense, LATTE is a general, flexible framework that can be instantiated in different ways to address specific setups. We describe the principles used in LATTE’s design (Section 4.1), followed by a detailed description of LATTE (Section 4.2).

## 4.1 The Design Principles of LATTE

Lookahead policies (also called Monte Carlo search policies) maximize the expected sum of future one-step rewards over simulated trajectories. In the power trading problem, the one-step reward is defined by Equation 3.1. In this dissertation we avoid using periodic/signup/withdraw fees, since we believe their current implementation in the Power TAC simulator has some issues that need to be fixed, and therefore we ignore the  $psw()$  component of the reward, which corresponds to these fees. When ignoring  $psw()$ , the reward becomes a function of eight variables  $Q_t^{cons}$ ,  $p_t^{cons}$ ,  $Q_t^{prod}$ ,  $p_t^{prod}$ ,  $Q_t^{ask}$ ,  $p_t^{ask}$ ,  $Q_t^{bid}$ ,  $p_t^{bid}$ , as seen in Equation 3.1. LATTE utilizes connections between these variables to address challenges (i) and (ii) and implement an efficient lookahead policy.

We address challenge (i) (of sampling actions efficiently) differently in each market. In the tariff market, LATTE samples the tariff action space in a reasonable region around the best existing tariffs, considering only a single tariff publication at a time rather than subsets of actions. The sampling resolution is determined by a real-time constraint on computation and the time it takes to predict a single action effect. Here an action effect consists of updates of state variables  $\mathcal{T}, \mathcal{S}_{B_0}, \mathcal{Q}_{B_0}$  (tariff subscriptions and consumption/production under each tariff), which in turn determine  $Q_t^{cons}$ ,  $p_t^{cons}$ ,  $Q_t^{prod}$ , and  $p_t^{prod}$  over the simulated horizon.

To address (i) in the wholesale market, LATTE uses a hierarchical two-step approach, based on the observation that each one-step reward  $Q_t^{bid} p_t^{bid}$  (or  $Q_t^{ask} p_t^{ask}$ )

is a *return* (sum of rewards) of a 24-step bidding policy  $\pi(Q)$  for buying (selling) quantity  $Q$  for a minimal (maximal) price. First, LATTE assumes in its lookahead that  $\pi$  is fixed and treats  $Q$  as an “action” (the decision to be made). LATTE uses an “action”-effect predictor  $Q_t^{bid} \xrightarrow{\pi} p_t^{bid}$  (or  $Q_t^{ask} \xrightarrow{\pi} p_t^{ask}$ ) that can be queried with sampled energy quantities and predict their average prices in the wholesale market. Using this predictor, LATTE predicts the reward components  $Q_t^{ask}$ ,  $p_t^{ask}$ ,  $Q_t^{bid}$ , and  $p_t^{bid}$ . Second, LATTE separates the subproblem of approximating a cost-optimizing sequential bidding policy  $\pi(Q)$  for trading quantities  $Q$ . In our instantiations of LATTE, this subproblem is solved using a small MDP isolated from the full MDP.

To address challenge (ii) (efficiently combining actions) we start by observing that there is a functional relationship between energy quantity  $Q_t^*$  and price  $p_t^*$  of each action type  $* \in \{cons, prod, ask, bid\}$ :  $p_t^*$  is typically an increasing/decreasing function of  $Q_t^*$  when buying/selling energy respectively. Therefore, we can focus on controlling and combining the traded quantities  $Q_t^{cons}, Q_t^{prod}, Q_t^{ask}, Q_t^{bid}$ . In the tariff market, a broker has direct control only over prices (through tariff publication actions), and the traded quantities are determined indirectly as a function of the broker’s tariffs. In the wholesale market, a broker controls both quantity and price in its bids, so it is convenient to think of the broker as first fixing a desired quantity to trade and then optimizing the price for which it is traded.

The reason why actions need to be optimized in conjunction is the imbalance fee  $bal(I_{B_0,t})$  paid for any imbalance  $I_{B_0,t} = Q_t^{cons} - Q_t^{prod} + Q_t^{ask} - Q_t^{bid}$ , which typically dominates trading costs in the tariff and wholesale markets. In general, a broker can benefit from imbalance (when it helps to counter-balance other brokers’ imbalanced portfolios). Given a *balancing predictor* which predicts  $bal(I_{B_0,t})$  for different imbalance values  $I_{B_0,t}$ , a lookahead policy can sample different imbalance levels and predict their cost/benefit for the broker.<sup>1</sup>

---

<sup>1</sup>In our implementations we always aimed to constrain the imbalance to 0 both for computational efficiency and since any imbalance does not seem beneficial in the current simulator.

To achieve a desired imbalance level, a broker needs to adjust the difference between the quantities it procures and sells. To do that efficiently, we take advantage of the temporal structure of the trading problem (Figure 3.1). Since tariffs result in repeated energy delivery and are published at low frequency, a single publication can (stochastically) determine  $Q_t^{cons}, Q_t^{prod}$  throughout the horizon. Since wholesale actions are for one-time delivery, and happen at high frequency, they can adapt the traded amounts more flexibly than tariff publications, separately for each timeslot. These properties gives rise to the following efficient combination of actions. First, given a sampled candidate tariff, predict the resulting quantities  $Q_t^{cons}, p_t^{cons}, Q_t^{prod}, p_t^{prod}$  throughout the lookahead horizon. Second, use these predicted quantities and the desired imbalance level to determine the (net) amount needed in the wholesale market in each future timeslot  $Q_t^{bid} - Q_t^{ask}$ . Third, sample combinations of  $Q_t^{bid}, Q_t^{ask}$  and predict their corresponding prices  $p_t^{bid}, p_t^{ask}$ . Fourth, combine the quantities  $Q_t^{cons}, p_t^{cons}, Q_t^{prod}, p_t^{prod}, Q_t^{bid}, Q_t^{ask}, p_t^{bid}, p_t^{ask}$  to compute the predicted *utility* (sum of rewards over the horizon) of this combination of actions.

LATTE uses this procedure to combine actions in its efficient lookahead policy, which repeatedly samples candidate tariffs and combines them with wholesale actions as described, predicts the expected utility of each combination, and executes the combination that maximizes the predicted utility. We note that such a lookahead policy can be viewed as a best-response strategy which does not take into account game-theoretic considerations: it assumes that the set of existing tariffs in the market remains fixed throughout the horizon, and does not try to predict opponents' responses to publishing a new tariff. Due to the complexity of opponent modeling in our domain, we leave it as an avenue for future work. We compensate for the lack of opponent modeling by frequently replanning, i.e. by executing LATTE frequently.

## 4.2 The LATTE Algorithm

This section presents the details of the LATTE algorithm. Figure 4.1 provides a visual overview of LATTE. A summary of the approach is as follows. When the broker agent receives the most recent observations from the environment, it executes LATTE. LATTE is a lookahead policy that generates and evaluates candidate action combinations (box 1), and executes the combination that is predicted to maximize utility in the tariff and wholesale markets (box 2). To generate a combination of actions, LATTE generates a candidate tariff action, which can be tariff publication, revocation or no-op (box 1.1); next, it predicts its effect on subscriptions and demand/production (box 1.2); next, it generates corresponding wholesale energy quantities to trade (box 1.4); next, it predicts the payments for trading these wholesale quantities (box 1.5); next, it combines these predictions to predict a utility over the future horizon (box 1.6). This computation flow can be augmented with an optional internal loop of generating and evaluating combinations with different imbalance levels (see box 1.3), which affect the quantities traded in the wholesale market. This step is currently unused in our implementation (See Section 4.1), but may become important in future versions of the simulator.

Note how action generation is interleaved with action-effect predictions: boxes 1.1 and 1.4 are action generation stages, while boxes 1.2 and 1.5 are action-effect prediction stages. Interleaving actions in this way allows for utilizing the temporal structure of the problem and constraining the search space, by taking into account the predicted tariff market quantities (and optionally the desired imbalance) when generating wholesale actions.

In box (1), the broker agent treats wholesale actions as quantities to trade (step (i) in the hierarchical approach described in Section 4.1). In box (2), wholesale quantities are converted into bids using a wholesale bidding policy  $\pi$  (step (ii) in the hierarchical approach).

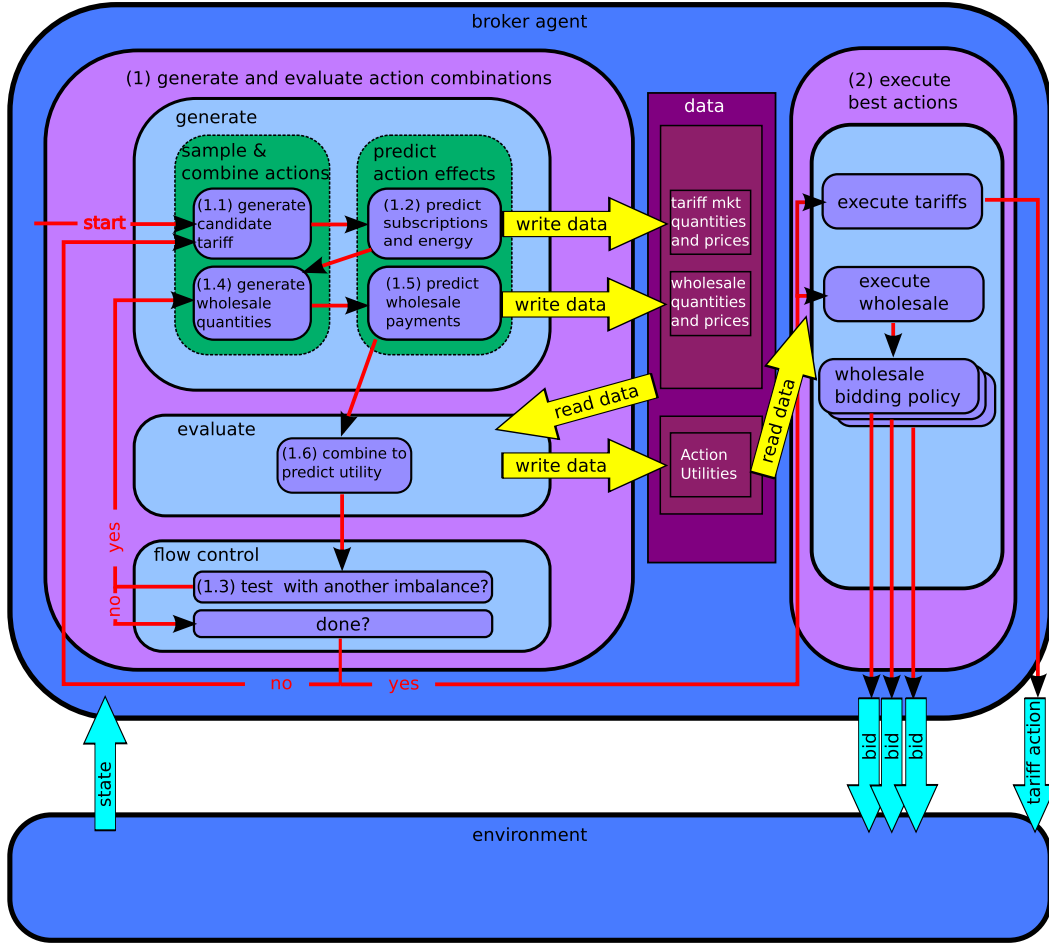


Figure 4.1: **The LATTE algorithm.** Thin (red) arrows indicate the computation flow. Rounded boxes indicates computation stages, which may be nested. Wide arrows indicate data read/write (yellow) and communication with the environment (light blue).

Algorithm 1 presents the pseudo-code of LATTE. Numbered comments correspond to computation states in Figure 4.1. Abstract functions that must be instantiated in specific implementations of LATTE are underlined. In particular, different implementations of these functions are specified in Chapters 5, 6, and 7 of this dissertation. LATTE has two phases. In phase 1 (lines 2-24) it generates action combinations and predicts their utilities. In phase 2 (lines 25-31) it executes the

best combination of actions in the tariff and wholesale markets.

In phase 1, if there is still time and phase 1 has not finished (line 3), **LATTE** generates and evaluates a combination of actions. First, a candidate tariff action is generated (line 5). This action is one of: publishing a new consumption or production tariff, revoking an existing tariff, and a no-op. Next, the tariff action effect is predicted, specifically the changes in tariff subscriptions and consumption/production under the new subscriptions. These predictions are used for predicting the total quantities and prices traded by the broker with its customers  $Q_t^{cons}, p_t^{cons}, Q_t^{prod}, p_t^{prod}$  and the quantities traded by the broker's competitors  $Q_t^{\oplus cons}, Q_t^{\oplus prod}$  (line 7).<sup>2</sup> Next, there is an optional step of trying to benefit from imbalance, by sampling imbalance quantities (line 9), and aiming for the sampled imbalances (lines 11, 13). A default implementation of this step uses only zero imbalance. Lines 10-20 generate wholesale actions and predict their effect throughout the simulated horizon. Lines 13-20 describe a degenerate implementation which does not separate wholesale buying and selling quantities in its optimization, but rather uses just one of them at a time.<sup>3</sup> Tariff market predictions and the desired imbalance determine the amount to trade in the wholesale market (line 13). Using the quantities it needs to trade in the wholesale market and the quantities predicted to be traded by competitors in the wholesale market (line 14), **LATTE** predicts the price of the traded amounts in the wholesale market (line 16). The reason it uses the amount traded by competitors is that wholesale prices are typically a function of the *total* amount traded by all brokers. Lines 17-20 assign values to one pair of variables based on whether the traded quantity is procured or sold. Next, **LATTE** combines all predictions into a utility prediction for the current combination of actions and

---

<sup>2</sup>We are using the symbol  $\oplus$  throughout to indicate the competitors' values of a variable.

<sup>3</sup>In the current simulator wholesale selling opportunities are much less frequent than wholesale buying opportunities: wholesale selling typically happens when customers leave a broker for competitors and the broker needs to sell any excess it had procured. Optimizing both wholesale buying and selling quantities might become important in future versions of the simulator, and could be addressed as we described in Section 4.1

---

**Algorithm 1** The LATTE Algorithm
 

---

▷ The main algorithm executed by TACTEX. Abstract functions are denoted as FunctionName.  
 1: **function** LATTE  
 2:   ▷ (1) Sample combinations of actions in tariff/wholesale markets and predict their expected utility  
 3:   **while** not Done() **do**  
 4:     ▷ (1.1) Generate candidate tariff  
 5:      $a^\tau \leftarrow \text{ComputeNextCandidateTariffAction}()$   
 6:     ▷ (1.2) Predict quantities/prices based on subscriptions and consumption/production  
 7:      $\{(Q_t^{cons}, p_t^{cons}, Q_t^{\oplus cons}, Q_t^{prod}, p_t^{prod}, Q_t^{\oplus prod}) | t = +1, \dots, +H\} \leftarrow \text{PredictTariffEffects}(a^\tau)$   
 8:     ▷ (1.3) Sample candidate imbalance levels  
 9:     **for**  $I \in \text{SampleImbalanceLevels}()$  **do**  
 10:       **for**  $t \in \{+1, \dots, +H\}$  **do**  
 11:          $I_{B_0,t} \leftarrow I$   
 12:         ▷ (1.4) Compute needed energy. ( $Q_t^{wholesale}$  is net wholesale quantity  $Q_t^{bid} - Q_t^{ask}$ ).  
 13:          $Q_t^{wholesale} \leftarrow Q_t^{cons} - Q_t^{prod} - I_{B_0,t}$    ▷ since  $I_{B_0,t} = Q_t^{cons} - Q_t^{prod} + Q_t^{ask} - Q_t^{bid}$   
 14:          $Q_t^{\oplus wholesale} \leftarrow \text{PredictCompetitorWholesaleQuantity}(Q_t^{\oplus cons}, Q_t^{\oplus prod})$   
 15:         ▷ (1.5) predict costs of procuring needed energy  
 16:          $p_t^{wholesale} \leftarrow \text{PredictWholesalePrice}(Q_t^{wholesale}, Q_t^{\oplus wholesale}, t)$   
 17:         ▷ Degenerate implementation: either buy or sell in the wholesale market, but not both  
 18:         **if**  $Q_t^{wholesale} > 0$  **then**  
 19:            $Q_t^{ask} \leftarrow 0, p_t^{ask} \leftarrow 0, Q_t^{bid} \leftarrow Q_t^{wholesale}, p_t^{bid} \leftarrow p_t^{wholesale}$   
 20:         **else**  
 21:            $Q_t^{bid} \leftarrow 0, p_t^{bid} \leftarrow 0, Q_t^{ask} \leftarrow Q_t^{wholesale}, p_t^{ask} \leftarrow p_t^{wholesale}$   
 22:         ▷ Record current action combination  
 23:          $a^\omega \leftarrow \langle Q_{+1}^{bid}, Q_{+1}^{ask}, Q_{+2}^{bid}, Q_{+2}^{ask}, \dots, Q_{+H}^{bid}, Q_{+H}^{ask} \rangle$   
 24:          $a \leftarrow a^\tau \cup a^\omega$   
 25:         ▷ (1.6) Combine predictions to an expected future utility  
 26:          $\text{utilities}[a] \leftarrow \sum_{t=+1}^{+H} \text{Reward}(Q_t^{cons}, p_t^{cons}, Q_t^{prod}, p_t^{prod}, Q_t^{ask}, p_t^{ask}, Q_t^{bid}, p_t^{bid}, \text{bal}(I_{B_0,t}), a^\tau)$   
 27:         ▷ (2) Execute the combination of actions in tariff/wholesale markets that maximizes predicted utility  
 28:          $a^* \leftarrow \arg \max_a \text{utilities}[a]$   
 29:         ExecuteActionInTariffMarket( $a^{*\tau}$ )   ▷ tariff market actions  
 30:          $\langle Q_{+1}^{bid}, Q_{+1}^{ask}, Q_{+2}^{bid}, Q_{+2}^{ask}, \dots, Q_{+H}^{bid}, Q_{+H}^{ask} \rangle \leftarrow a^{*\omega}$    ▷ wholesale market actions  
 31:         **for**  $t$  in  $+1, +2, \dots$  **do**  
 32:           WholesaleBiddingPolicy( $Q_t^{bid}, t$ )  
 33:           WholesaleBiddingPolicy( $Q_t^{ask}, t$ )

---

 ▷ Function implementations common to all instantiations:
 

---

▷ Reward function from Equation 3.1  
 32: **function** Reward( $Q_t^{cons}, p_t^{cons}, Q_t^{prod}, p_t^{prod}, Q_t^{ask}, p_t^{ask}, Q_t^{bid}, p_t^{bid}, \text{bal}(I_{B_0,t}), a^\tau$ )  
 33:    $r_t \leftarrow Q_t^{cons} p_t^{cons} - Q_t^{prod} p_t^{prod} + Q_t^{ask} p_t^{ask} - Q_t^{bid} p_t^{bid} +$   
 34:    $\text{bal}(I_{B_0,t}) - \text{dist}(Q_t^{cons}, Q_t^{prod}) - \text{pub}(a^\tau) - \text{rev}(a^\tau)$   
 35:   **return**  $r_t$

---

▷ Default implementation assuming competitors will be balanced (buy exactly what they need)  
 36: **function** PredictCompetitorWholesaleQuantity( $Q_t^{\oplus cons}, Q_t^{\oplus prod}$ )  
 37:    $Q_t^{\oplus wholesale} \leftarrow Q_t^{\oplus cons} - Q_t^{\oplus prod}$   
 38:   **return**  $Q_t^{\oplus wholesale}$

---



records the mapping from actions to utility (lines 21-24).

In phase 2, LATTE selects the action combination that maximizes the predicted utility (line 26), executes the tariff action in the tariff market (line 27)<sup>4</sup>, and trades the desired amounts in the wholesale market (lines 28-31) using a wholesale bidding policy that aims to minimize their costs.<sup>5</sup>

There are two function implementations that are common to all our instantiations of LATTE: the `Reward()` function and the `PredictCompetitorWholesaleQuantity()` function. The `Reward()` function sums the predicted reward components into a complete one-step reward (lines 32-34). The `PredictCompetitorWholesaleQuantity()` predicts that competitors will trade in the wholesale market the quantity that will result in zero imbalance in their portfolio (lines 35-37). This assumption does not necessarily hold, and therefore more sophisticated implementations of this function could be explored in future work.

### 4.3 Chapter Summary

In this chapter we introduced LATTE, which implements an approximate solution to the energy trading MDP in the form of an efficient lookahead policy. Based on Powell’s characterization of lookahead policies [80], LATTE can be classified as a rolling-horizon, Monte-Carlo search policy. We have described the principles used in the design of LATTE as well as the algorithm itself, both graphically and in pseudo-code. LATTE contains abstract methods that can be implemented in different ways in dif-

---

<sup>4</sup>In practice, while LATTE is invoked at every timeslot, our implementation allows it to publish tariffs only once every 6 hours in the first four days (which is the minimal publication interval in Power TAC), and thereafter only once every 24 hours. This implementation is equivalent to running LATTE with an implementation of `ComputeNextCandidateTariffAction` that returns a no-op in every timeslot in which tariff publications are not allowed. The reason we artificially limit tariff publications is that our subscription predictors ignore customers’ inertia and this may lead to over publishing in some cases.

<sup>5</sup>Since the wholesale market is a day-ahead market, `WholesaleBiddingPolicy` can only be called for the next 24 timeslots, even if the horizon  $H$  is longer than 24 timeslots. However, since LATTE runs at every timeslot, eventually it bids for every timeslot in the horizon.

ferent setups. The following chapters will describe specific instantiations of **LATTE** that were used effectively in practice. Specifically, these implementations consistently achieved state-of-the-art performance, both in competitions and in controlled experiments.

## Chapter 5

# TacTex-13: A Champion Adaptive Power Trading Agent

This chapter introduces TacTex-13, the champion broker agent from the inaugural Power TAC competition in 2013. TacTex-13 is a complete, fully implemented broker agent that instantiates the LATTE algorithm with components that learn online to predict the tariff and wholesale market action effects, and to bid in the wholesale market. This chapter describes the constituent components of TacTex-13's instantiation of LATTE (Section 5.1), and examines TacTex-13's success through analysis of competition results and subsequent controlled experiments (Section 5.2). TacTex-13's binary is publicly available, along with source code of more recent versions, which can be configured through a text file with TacTex-13's components (see Appendix A).

---

<sup>0</sup>This chapter is based on a published conference paper [111] that I wrote with Professor Peter Stone. Author contributions were as follows: I was a Ph.D. student and did the complete implementation, and Peter was my advisor and collaborated with me on deciding on research directions and analyzing and interpreting results.

## 5.1 TacTex-13 Description

TacTex-13 uses a restricted instantiation of LATTE that procures power in the whole-sale market, sells power in the tariff market, and aims for zero imbalance. To sell power in the tariff market, TacTex-13 uses only fixed-rate consumption tariffs and does not use tariff-revoke actions. This instantiation restricts and simplifies LATTE in multiple ways, and results in several reward components being zero throughout the LATTE algorithm, as follows (see the complete reward specification in Algorithm 1, lines 32-34):

- Since TacTex-13 does not use production tariffs,  $Q_t^{prod} = 0$  and  $p_t^{prod} = 0$ .
- Since TacTex-13 aims for zero imbalance,  $I_{B_0,t} = 0$  and  $bal(I_{B_0,t}) = 0$ .
- Since  $Q_t^{prod} = 0$  and  $I_{B_0,t} = 0$ , we get that  $Q_t^{bid} = Q_t^{wholesale} = Q_t^{cons}$ , and that  $Q_t^{ask} = 0$  and  $p_t^{ask} = 0$  (based on lines 13, 17, 18 of Algorithm 1).
- Since TacTex-13 never uses tariff-revoke actions,  $rev(a^\tau)$  is always 0.

Next, we describe how TacTex-13 instantiates the abstract functions defined in LATTE’s phase 1 (Section 5.1.1) and phase 2 (Section 5.1.2).

### 5.1.1 TacTex-13’s Instantiation of LATTE’s Phase 1

#### Instantiating `ComputeNextCandidateTariffAction()`

TacTex-13 instantiates the `ComputeNextCandidateTariffAction()` function as follows. The first call to this function generates a set of candidate fixed-rate tariffs and returns the first candidate; each subsequent call returns the next candidate in the set. The set of candidate tariffs is generated as 100 fixed-rate tariffs with prices that are equally spaced in a range that contains the currently best published tariff prices. More specifically, let  $rng := [p_1, p_2]$  be the price range that lies between the rate

of the best tariff published by TacTex-13 and the (average) rate of the best tariff published by its competitors; then the candidate tariffs are generated in the range  $[0.8p_1, 1.2p_2]$ .

The numbers 0.8 and 1.2 were initially chosen heuristically, to allow for price reduction or increase of up to 20%. Later, informal experimentation showed that perturbing these values did not generally improve performance. Specifically we observed that: (i) perturbing the upper bound of 1.2 by 0.1 did not significantly affect performance; (ii) perturbing the lower bound of 0.8 by 0.1 affected performance only when playing against a *small number of cooperative* competitors (which do not reduce prices aggressively): using a value of 0.9 or 0.95 instead of 0.8 resulted in milder price reductions, and therefore in higher scores (profits) for all brokers, but reduced the broker’s market share and thus increased the risk of losing the game; (iii) perturbing the lower bound of 0.8 to 0.7 did not significantly improve performance; and (iv) perturbing the lower bound to 0.3 significantly reduced performance in games where initial wholesale cost predictions under-estimated the actual wholesale costs, since the broker reduced prices more than it should have, and its profit decreased significantly.

This instantiation of `ComputeNextCandidateTariffAction()` has several benefits for the broker. First, it simplifies the optimization: LATTE searches solely over one type of tariff action, namely fixed-rate tariff publications, and therefore optimizes only one selling-price rather than a separate price for each future timeslot. Moreover, fixed-rate tariffs reduce stochasticity in action effects, since they do not encourage customers to shift consumption, thus making customers’ behavior more predictable. Second, customers in Power TAC prefer tariffs that do not cause any discomfort (i.e. consumption shifts), so using fixed-rate tariffs makes it easier for TacTex-13 to acquire customers.<sup>1</sup>

---

<sup>1</sup>Clearly, shifting consumption to follow energy supply is one of the primary goals of future smart grids. Tariffs that encourage that are aligned with this goal. Chapter 7 investigates the usage of

### Instantiating **PredictTariffEffects**( $a^\tau$ )

TacTex-13’s instantiation of **PredictTariffEffects** is presented in Algorithm 2. This instantiation predicts only the  $Q_t^{cons}$  and  $p_t^{cons}$  variables, since the other predicted variables from line 7 of LATTE (Algorithm 1) are unused by TacTex-13. The length of the horizon  $H$  over which these variables are predicted is one week ( $7 \times 24 = 168$  timeslots). We chose this length as a trade-off between shorter horizons, which might not capture weekly consumption/production patterns, and longer horizons which present higher uncertainty and require more computation.

At the core of Algorithm 2 lies the problem of estimating the demand resulting from a tariff publication. We decompose this problem into two problems: (1) predicting the resulting customer migration between tariffs (line 4), and (2) predicting the demand of each of the customers over the lookahead horizon (line 5). The former is detailed below in Algorithm 3. The latter is addressed by maintaining records with average past demand for each customer, in each of the 168 weekly timeslots, and using these records to predict future demand. This implementation ignores weather conditions and uses only the time of week to predict consumption. It also ignores demand elasticity and does not consider customers’ adjustment of demand based on the tariff they are subscribed to. While these simplifications may fail to capture complex patterns in the data, they help the broker to learn online to predict demand with little data. This tradeoff between the complexity of predictive models and the amount of data needed for learning them is generally a key tradeoff in machine learning (known as the *bias-variance tradeoff*), and specifically a key tradeoff in learning to predict customer demand. The small amounts of data available for online learning in Power TAC were the reason we used simplified predictive models with strong generalization.

Using the information gathered in lines 4-5, the total demand and income are such tariffs by self-interested brokers.

computed by summing over all customer-tariff pairs (line 6-11). Lines 12-14 convert the total demand and income into the  $Q_t^{cons}$  and  $p_t^{cons}$  variables which are returned as the predicted action effects of the tariff publication  $a^\tau$ .

---

**Algorithm 2** TacTex-13::PredictTariffEffects( $a^\tau$ )

---

```

    ▷ Initializing variables
1:  $tariff \leftarrow a^\tau$  ▷ Renaming for readability
2:  $totalDemand[1, \dots, 168] \leftarrow [0, \dots, 0]$ 
3:  $totalIncome[1, \dots, 168] \leftarrow [0, \dots, 0]$ 

    ▷ Predicting effects
4:  $subscriptions[:, :] \leftarrow \text{PredictChangeInSubscriptions}(tariff)$ 
5:  $demandProfiles[:, :] \leftarrow \text{PredictCustomerDemandProfiles}()$ 
6: for  $cust \in customerModels$  do
7:   for  $tariff \in \{tariff \cup existingTariffs\}$  do
8:      $n \leftarrow subscriptions[cust, tariff]$ 
9:      $customerDemand[1 \dots 168] \leftarrow n \times demandProfiles[cust]$ 
10:     $totalDemand[1 \dots 168] \leftarrow totalDemand[1 \dots 168] + customerDemand[1 \dots 168]$ 
11:     $totalIncome[1 \dots 168] \leftarrow totalIncome[1 \dots 168] + customerDemand[1 \dots 168] \cdot tariff.rate()$ 
12: for  $t \in \{1, 2, \dots, 168\}$  do
13:    $Q_t^{cons} \leftarrow totalDemand[t]$ 
14:    $p_t^{cons} \leftarrow \frac{totalIncome[t]}{totalDemand[t]}$ 
return  $\{(Q_t^{cons}, p_t^{cons}, 0, 0, 0, 0) | t = +1, \dots, +168\}$ 

```

---

Algorithm 3 describes the function `PredictChangeInSubscriptions()`, which is used in line 4 of Algorithm 2. `PredictChangeInSubscriptions()` predicts the changes in subscriptions as a result of a new tariff publication. It starts with predicting all customer demand profiles (line 1), similarly to line 5 of Algorithm 2.<sup>2</sup> TacTex-13 predicts the change in subscriptions separately for each customer (line 2). Recall that Power TAC customers represent whole populations and can subscribe subsets of their population to different tariffs. For each tariff, the predicted weekly demand of a single member of the population (line 3) is multiplied by this tariff's rate to compute the expected weekly charge for a single member under this tariff (line 6, implemented similarly to the right term in line 11 of Algorithm 2); then it is paired with the number of individuals currently subscribed to this tariff to form a pair  $\langle charge, numSubs \rangle$  that is added to a set of such pairs (line 8). The pairs  $\langle charge, numSubs \rangle$  are used as a training set for a supervised learning al-

---

<sup>2</sup>In practice, instead of repeating this function call in each of the algorithms, we cache results and use them.

gorithm, specifically Locally Weighted Linear Regression (LWR), that predicts the subscribed-population size for the candidate tariff based on its expected charge (lines 9-12). LWR (see, e.g. [3]) was chosen since, being non-parametric, it requires very few assumptions about the representation of the predicted function (the customer preference function).<sup>3</sup>

Since new subscribers to the candidate tariff must migrate from other tariffs (published by either TacTex-13 or its competitors), TacTex-13 uses a normalization step after which all tariff subscriptions are scaled proportionally so that the total number of predicted subscriptions equals the customer’s population size (line 13). Typically, part of the population is subscribed to competitors. In line 10, *charge2subs* must represent the entire customer population (e.g. to be able to predict migration of customers from competitors to TacTex-13), even though subscriptions to competitors are unobservable by TacTex-13. To represent the complete population, a dummy subscription is added to *charge2subs*, which assigns the portion of the population that is not subscribed to TacTex-13’s tariffs to the best competitor tariff. Finally, all the predicted subscriptions for this customer are added to a map (lines 14-15) that is returned by the algorithm.

### Instantiating **SampleImbalanceLevels()**

TacTex-13 always aims for zero imbalance and therefore uses a degenerate instantiation of **SampleImbalanceLevels()** that returns the set  $\{0\}$ .

### Instantiating **PredictWholesalePrice**( $Q_t^{wholesale}, Q_t^{\oplus wholesale}, t$ )

TacTex-13 predicts energy costs similarly to how it predicts customer demand profiles: it maintains a record of past average costs in each of the 168 weekly times-

---

<sup>3</sup>This implementation ignores customer inertia, i.e. the tendency to stay with the current tariff. Therefore, it predicts the *eventual* subscriptions if no further action is taken in the market by any broker.



---

**Algorithm 3** TacTex-13::PredictChangeInSubscriptions(*candidateTariff*)

---

```
1: demandProfiles[.]  $\leftarrow$  PredictCustomerDemandProfiles()
2: for cust  $\in$  customers do
3:   customerDemand[1...168]  $\leftarrow$  demandProfiles[cust]
4:   charge2subs  $\leftarrow$  {}
5:   for tariff  $\in$  existingTariffs do
6:     charge  $\leftarrow$  ExpectedTariffCharge(customerDemand, tariff)
7:     numSubs  $\leftarrow$  currentSubscriptions[cust, tariff]
8:     charge2subs  $\leftarrow$  charge2subs  $\cup$  {charge, numSubs}
9:   charge  $\leftarrow$  ExpectedTariffCharge(customerDemand, candidateTariff)
10:  trainingSet  $\leftarrow$  charge2subs
11:  numSubs  $\leftarrow$  PredictWithLWR(trainingSet, charge)
12:  charge2subs  $\leftarrow$  charge2subs  $\cup$  {charge, numSubs}
13:  charge2subs  $\leftarrow$  Normalize(charge2subs)
14:  for tariff  $\in$  candidateTariff  $\cup$  existingTariffs do
15:    predSubs[cust, tariff]  $\leftarrow$  ExtractSubscriptions(charge2subs, tariff)
return predSubs
```

---

lots and use it to predict future costs based on the time of week  $t$ . Doing so assumes that energy cost is independent of the quantity predicted to be procured by TacTex-13 ( $Q_t^{wholesale}$ ) or its competitors ( $Q_t^{\oplus wholesale}$ ), and that energy costs in a given weekday/time-of-day combination are similar between different weeks. These assumptions hold when customer weekly consumption patterns are similar between different weeks (which can happen when weather is similar and when brokers use only fixed rate tariffs which discourage consumption shifting), and when the energy quantity procured in the wholesale market by all brokers is roughly equal to the quantity consumed by all customers in every timeslot. These assumptions roughly held in the Power TAC 2013 finals, where brokers mostly used fixed-rate tariffs, and procured most of their energy in the wholesale market.

### 5.1.2 TacTex-13's Instantiation of LATTE's Phase 2

In phase 2, LATTE decides on the quantities to trade in the wholesale market for future *target timeslots*  $t$ . These quantities ( $Q_t^{bid}$  or  $Q_t^{ask}$ ) are denoted here as  $Q_t$ . TacTex-13 sends these quantities to a wholesale bidding policy that is

encapsulated in the abstract function  $\text{WholesaleBiddingPolicy}(Q_t, t)$ . This section describes **TacTex-13**'s instantiation of  $\text{WholesaleBiddingPolicy}(Q_t, t)$ .

In the wholesale market **TacTex-13** primarily procures energy (to satisfy the demand of its customers), and only sells energy when it predicts that future demand would be smaller than the quantity it had already procured (e.g. as a result of customers migrating to competitors). Since **TacTex-13** mostly procures energy in the wholesale market, its wholesale bidding policy is designed to minimize the cost of procuring a quantity  $Q_t$ . When **TacTex-13** needs to sell energy in the wholesale market, it uses a default randomized policy that reduces its selling-price offers as time gets closer to the target timeslot  $t$ .

To minimize the energy costs **TacTex-13** needs to (1) minimize the rates for which it procures energy in the wholesale market, and (2) minimize its imbalance costs, by satisfying the future demand as accurately as possible. To do the latter, it must (2.1) have accurate predictions of future demand (provided by **LATTE**), and (2.2) be able to procure all the energy predicted to be demanded. For notational convenience, we will denote here a target timeslot as  $t_{tar}$  and a general timeslot as  $t$ . The actions that affect the energy cost for a target timeslot  $t_{tar}$  are the 24 bidding (or not-bidding) actions in each of the 24 preceding timeslots,  $(t_{tar} - 24, \dots, t_{tar} - 1)$ , which thus comprise a sequential bidding process with 24 steps. Thus, at each timeslot  $t$ , **TacTex-13** executes, respectively, steps  $1, 2, \dots, 24$  of 24 independent bidding processes for timeslots  $t + 24, \dots, t + 1$ .

**TacTex-13**'s wholesale market bidding policy uses a modified version of Tesauro's bidding algorithm [105]. We model the sequential bidding process as a Markov Decision Process (MDP) [82] in a specific way that allows for computational efficiency, and more importantly in the competitive environment that **TacTex-13** operates in, it allows for high reuse of data, and thus quick online learning with little data. The bidding MDP is isolated from the complete power trading MDP, by using a

subset of state variables and actions of the complete MDP to define the bidding MDP. TacTex-13’s bidding MDP is defined next, followed by the rationale behind its design:

- **States:**  $s \in \{0, 1, \dots, 24, success\}$ ,  $s_0 := 24$
- **Actions:** limit-price  $\in \mathbb{R}$
- **Transition:** a state  $s \in \{1, \dots, 24\}$  transitions to one of two states. If a bid is partially or fully cleared, it transitions to the terminal state *success*. Otherwise, a state  $s$  transitions to state  $s - 1$ . The clearing (i.e. transition) probability  $\rho_{cleared}(s, \text{limit-price})$  is initially unknown.
- **Reward:** In state  $s = 0$ , the reward is the balancing-price per energy unit. In states  $s \in \{1, \dots, 24\}$ , the reward is 0. In state *success*, the reward is the limit-price of the successful bid. Both balancing-price and limit-price are taken as negative, so maximizing the reward results in minimizing costs. balancing-price is initially unknown.
- **Terminal States:**  $\{0, success\}$

In a sequential bidding process for a target timeslot, the broker actions are bids of the form `bid(energy-amount,limit-price)`. Tesauro’s bidding MDP uses these actions as the MDP actions. However, we excluded energy-amount from the decision making; it is always set to the difference between predicted demand  $Q_{t_{tar}}$  and the energy that is already procured for the target timeslot (TacTex-13’s market-position, represented in the complete MDP’s state variable  $\mathcal{M}_{B_0, t_{tar}}$ ).

The solution to our MDP is a sequential bidding policy that minimizes the expected energy unit-cost for the next fraction of the procured quantity. Note that there is a transition to a terminal state *success* even in cases where the bid is partially cleared. One implication of excluding energy-amount from the MDP’s state

and action representations is that every sequential bidding process executes over the *same sequence of states*, thus allowing for computational and data efficiency, as seen next.

Since the MDP is acyclic (linear), solving it requires one back-sweep, starting from state 0 back to state 24, applying the following backup operator to compute a value function:

$$V(s) = \begin{cases} \text{balancing-price} & \text{if } s = 0 \\ \min_{\text{limit-price}} \{ \rho_{\text{cleared}} \times \text{limit-price} + (1 - \rho_{\text{cleared}}) \times V(s-1) \} & \text{if } 1 \leq s \leq 24 \end{cases}$$

The MDP’s solution determines an optimal (over this MDP which approximates the real MDP) limit-price for each of the 24 states. Using our MDP model, **TacTex-13** is always in states  $1, \dots, 24$  of 24 concurrent bidding processes. Therefore, **TacTex-13** solves the MDP once per timeslot, and submits the 24 optimal limit-prices to the 24 auctions.

Before solving this MDP, **TacTex-13** needs to learn the MDP’s unknown quantities, namely the expected **balancing-price** at  $s = 0$  and the transition function  $\rho_{\text{cleared}}$ . **TacTex-13** learns the transition function from past data by recording, for each state  $s \in \{1, \dots, 24\}$ , the wholesale trades executed in  $s$  into a set  $P_s$  (here  $P$  stands for density). Each trade has the form (clearing-price, cleared-energy-amount) ( $p^c$  and  $q^c$  from the complete MDP’s auction results  $\mathcal{A}_{B_0}$ ). The set  $P_s$  is treated as a non-parametric density estimation and a transition probability is estimated from it as  $\rho_{\text{cleared}}(s, \text{limit-price}) := \frac{\sum_{tr \in \text{trades}[s], tr.\text{clearing-price} < \text{limit-price}} tr.\text{cleared-energy-amount}}{\sum_{tr \in \text{trades}[s]} tr.\text{cleared-energy-amount}}$ . To estimate the mean **balancing-price**, **TacTex-13** similarly maintains a set  $P_0$  of past balancing data. Since every bidding MDP executes over the same sequence of states  $s \in \{0, \dots, 24\}$ , every trade executed in state  $s$  can be used by all future bidding processes as a part of  $P_s$ . Thus, our state representation allows **TacTex-13** to efficiently reuse data and thus speed-up learning. Clearly, our state representation relies on the assumption that time-to-target-timeslot is a dominant feature in determining

the transition function, i.e. the distribution of auction closing prices. Were that not the case, other features would need to be added to the MDP’s state.

TacTex-13’s bidding policy is summarized in Algorithm 4 which is TacTex-13’s main routine in the wholesale market, executed at every timeslot. It computes the needed energy for the coming 24 timeslots using demand-predictions and market-positions (line 1), then adds the previous timeslot’s wholesale market trades and balancing information to the  $P_s$  sets (line 2). If not enough trades were recorded for each state (specifically fewer than 6), a randomized bidding policy is executed, otherwise the MDP-based bidding policy is executed (lines 3-7). The number 6 was chosen to trade off quick learning with reasonable density estimations.

---

**Algorithm 4** Online RL Wholesale Market Strategy

---

```

1: neededEnergy[1 . . . 24] = ComputeNeededEnergy()
2: densities[0 . . . 24]  $\leftarrow$  AddRecentTradesAndBalancing()
3: if HasEnoughData(densities) then
4:   limitPrices[1 . . . 24] = SolveMDP(densities)
5: else
6:   limitPrices[1 . . . 24] = RandomizedBiddingPolicy()
7: SubmitBids(neededEnergy[1 . . . 24], limitprices[1 . . . 24])

```

---

To summarize, TacTex-13 starts a game with no data and learns to bid online, while acting. Its estimates are refined during the game as it collects more data. At each timeslot, it solves the MDP with all the data collected so far. The result is an online reinforcement learning (RL) bidding algorithm that allows TacTex-13 to adapt and optimize its bidding policy to each game’s specific market conditions.

## 5.2 Results

This section examines the success of TacTex-13 through analysis of the competition and controlled experiments.

### 5.2.1 Competition Results: Power TAC 2013 Finals Analysis

The Power TAC 2013 finals were held in conjunction with the AAAI'13 conference. The qualifying competitors were 7 brokers developed by research groups from Europe and the USA. The competition included all possible combinations of 2-broker and 4-broker games (21 and 35 games respectively), and 4 7-broker games. Table 5.1 shows the final cumulative scores in each of the game sizes, the final z-scores in each of the game sizes, and competition totals. The final ranking is determined by the rightmost column, which sums the z-scores of all game sizes. TacTex-13 had both the highest total z-scores and the highest cumulative scores. In the 2-broker games TacTex-13 won all of its 6 games. In the 4-broker games, TacTex-13 won 15 out of the 16 games it completed successfully (TacTex-13 got disconnected from 4 games due to technical issues with the infrastructure we used). TacTex-13 did not win the 7-broker games despite having the largest volume of customers. Next, we analyze these results.

Table 5.1: Results of the Power TAC 2013 finals

Broker	Cumulative Scores				Z-Scores			
	7-broker	4-broker	2-broker	Total	7-broker	4-broker	2-broker	Total
TacTex-13	-705248	13493825	17853189	30641766	0.386	0.449	0.691	<b>1.526</b>
cwiBroker	647400	12197772	13476434	26321606	0.437	0.442	0.536	<b>1.415</b>
MLLBroker	8533	3305131	9482400	12796064	0.413	0.391	0.395	<b>1.199</b>
CrocodileAgent	-361939	1592764	7105236	8336061	0.399	0.381	0.311	<b>1.091</b>
AstonTAC	345300	5977354	5484780	11807435	0.425	0.406	0.254	<b>1.086</b>
Mertacor	-621040	1279380	4919087	5577427	0.389	0.380	0.234	<b>1.003</b>
INAOEBroker02	-76112159	-497131383	-70255037	-643498580	-2.449	-2.449	-2.421	<b>-7.319</b>

Figure 5.1 shows averages of the main components of the brokers' cash flow, for each of the game sizes.<sup>4</sup> Brokers are ordered based on their final ranking in the competition, from left to right. For each broker, the bars show (from left to right) its average (1) profit (2) income from consumption tariff subscribers (3) tariff publication fees (proportional the number of tariffs published) (4) wholesale market costs (5) balancing costs, and (6) energy distribution costs (proportional to the

<sup>4</sup>We excluded INAOEBroker; its large negative scores, caused by large tariff-publication fees, affected the readability of the plots.

amount of traded energy).

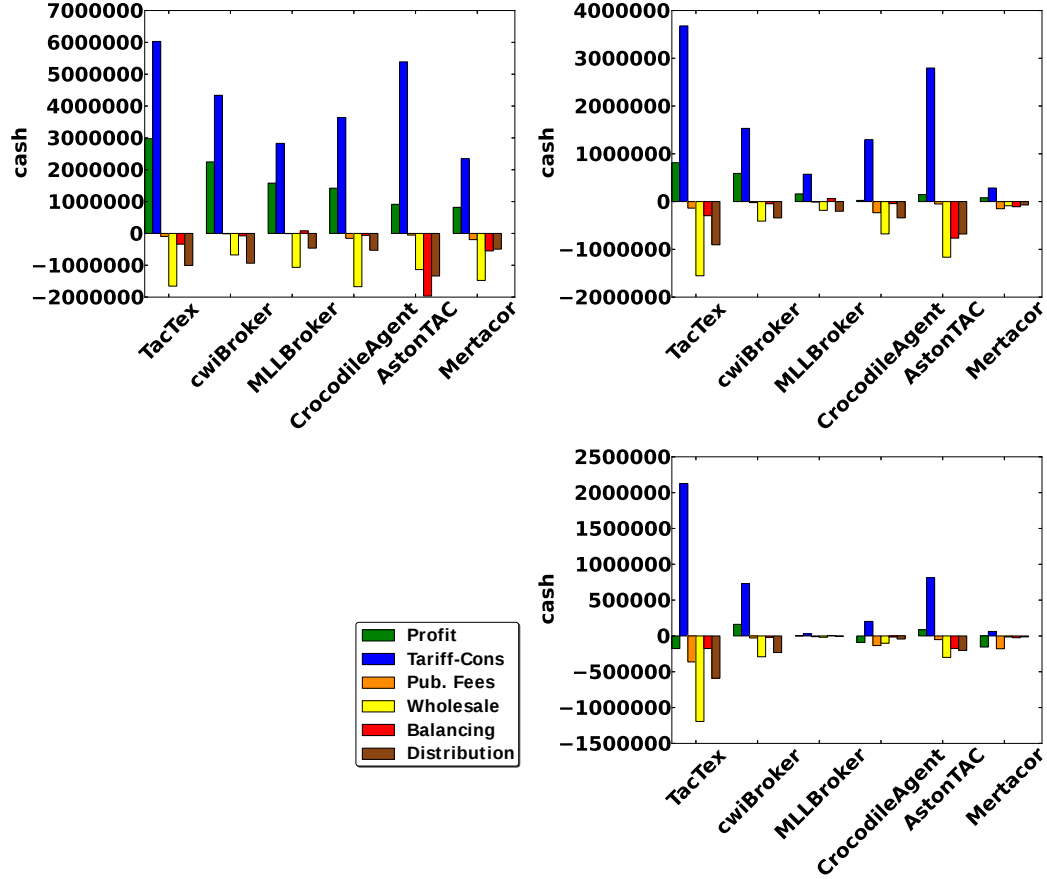


Figure 5.1: Power TAC 2013 finals: average income and costs in 2-broker (top-left), 4-broker (top-right), and 7-broker games (bottom-right)

At a high level, TacTex-13's wholesale market bidding policy and the tariff market strategy induced by LATTE were responsible for TacTex-13's success in the finals. The wholesale market policy maintained low-costs, while the tariff market strategy balanced its offered tariff prices with the resulting predicted demand to optimize profits given the costs achieved by the wholesale policy. More specifically, in the 2-broker games TacTex-13 made 32.4% and 88.2% more profits than the 2nd (cwi) and 3rd (MLL) place brokers while maintaining similar levels of income-to-costs

ratio (1.97), compared to cwi’s (2.07) and MLL’s (2.26). In Power TAC’s wholesale market, energy unit-cost is typically an increasing function of the procured amount. Despite that, TacTex-13 sold 50.5% and 72.5% more energy than cwi and MLL with a competitive cost-per-kWh (4.4 cents/kWh) compared to cwi’s and MLL’s (4.6, 3.1 cents/kWh)<sup>5</sup>. It can be seen that the majority of TacTex-13’s costs were spent on wholesale market procurement and (non-controllable) distribution fees. Therefore, TacTex-13’s low cost-per-kWh is attributed to its wholesale market policy. At the same time, given these costs, its tariff market strategy published tariffs at an average rate that is slightly lower than cwi’s and slightly higher than MLL’s (8.8, vs 9.5 and 7.1 cents/kWh), which resulted in 39.0% and 113.6% more income compared to cwi and MLL. In the 4-broker games, TacTex-13 traded 9% less energy comparing to the 2-broker games, while maintaining similar average wholesale market costs. Due to the stronger competition, TacTex-13’s income decreased by 61%, since its tariff market strategy recognized it had to reduce prices (by 66.6%) to maximize its profits. TacTex-13’s profits (and income) were higher by 38.1% (139.9%) and 404.5% (542.2%) compared to cwi’s and MLL’s, while its income-to-cost ratio decreased to 1.28 compared to 1.62 and 1.39 of cwi and MLL. In the 7-broker games, TacTex-13’s tariff strategy had to lower prices further, but also recognized a stopping point beyond which it did not decrease rates. However, due to an underestimation of the predicted costs, TacTex-13 ended up with losses despite having large customer volume and income.

### 5.2.2 Controlled Experiments

We performed controlled experiments to identify the contribution of each of TacTex-13’s major components. To do that, we generated test agents by disabling components of TacTex-13 and comparing the resulting performance. Specifically, agent

---

<sup>5</sup>Not shown in the figure.



U9\_MDP\_LWR was the full TacTex-13 agent. Agent U9\_MDP was generated from U9\_MDP\_LWR by removing the LWR-based customer-subscriptions prediction component and replacing it with linear interpolation and conservative extrapolation, thus modifying the implementation of `PredictTariffEffects()`. Agent U9 was generated from U9\_MDP by disabling TacTex-13’s MDP-based instantiation of the function `WholesaleBiddingPolicy()` and replacing it with a baseline, randomized policy that starts by trying lower buying prices and increasing them as time gets closer to target timeslot. Agent U1 was generated from U9 by publishing 1, instead of 9, initial sample tariffs for probing customer tariff subscriptions, used by Algorithm 3. Finally, a baseline agent B was generated from U1 by disabling LATTE’s lookahead-based tariff market strategy and replacing it with a strategy that reacted to competitor tariffs by publishing slightly better rates. Specifically, LATTE’s pseudo-code was degenerated by: (a) executing the code in the while loop (lines 3-20) only once, (b) using the reactive tariff strategy to instantiate the abstract function `ComputeNextCandidateTariffAction()`, (c) eliminating lines 23-24 which record utility, and (d) replacing line 26 with  $a^* \leftarrow a$ . The result was publishing tariffs based on the reactive tariff strategy, and determining wholesale quantities to procure using (the above modified implementation of) `PredictTariffEffects()` for these tariffs.

Table 5.2: Round-Robin ablation analysis.

	<b>B</b>	<b>U1</b>	<b>U9_MDP</b>
<b>U9_MDP_LWR</b>	1278.3 (43.2)	708.9 (35.6)	34.2 (23.2)
<b>U9_MDP</b>	966.4 (40.5)	592.6 (22.2)	
<b>U1</b>	547.4 (27.7))		

We compared the above agents in two groups of experiments. The first group was a 2-broker round-robin tournament between U9\_MDP\_LWR, U9\_MDP, U1 and B. The second group compared the performance of all versions in 4-broker games against a fixed set of opponents, composed of the 3 finalist broker binaries that were available to us: `cwiBroker`, `CrocodileAgent` and `Mertacor`. In all of our experiments,

Table 5.3: Ablation analysis using 3 finalist broker agents.

Broker	Cash	Broker	Cash
cwiBroker	340.9 (8.4)	cwiBroker	315.4 (9.3)
Mertacor	-276.2 (40.2)	<b>U1</b>	<b>135.3 (12.3)</b>
CrocodileAgent	-287.1 (14.5)	CrocodileAgent	-372.1 (17.0)
<b>B</b>	<b>-334.6 (8.0)</b>	Mertacor	-485.5 (28.1)

Broker	Cash	Broker	Cash
cwiBroker	316.2 (9.1)	<b>U9_MDP</b>	<b>389.9 (13.3)</b>
<b>U9</b>	<b>182.8 (12.4)</b>	cwiBroker	138.3 (8.7)
CrocodileAgent	-338.2 (17.0)	CrocodileAgent	-333.3 (17.0)
Mertacor	-476.6 (28.6)	Mertacor	-494.1 (29.6)

Broker	Cash
<b>U9_MDP_LWR</b>	<b>350.8 (13.3)</b>
cwiBroker	132.4 (9.0)
CrocodileAgent	-336.9 (17.3)
Mertacor	-566.1 (26.8)

each given combination of agents was tested over a fixed set of 200 full games. Each game took about 2 hours of real-time (about 60 days of simulated time), and was generated by loading a set of random-number seeds that initialized the random number generators of the simulation, and a weather data file that completely determined the simulated weather. We note that even after loading weather and seeds, there was still some randomness of unknown source in the simulation. Each weather file contained around 3 months of real-world weather, recorded in the default location simulated by Power TAC. We used 8 different weather files (each file used by 25 out of the 200 games), using the recording start dates of January, April, July, October of 2009 and 2010, thus covering a period of 2 years. The simulator version used in the experiments in this section is specified in Appendix A. The results of the first group of experiments are reported in Table 5.2. Each entry in the table is the mean score-difference (in 1000s) over the set of 200 games. The results of the second group of experiments is reported in Table 5.3. Each of the 5 two-column

sub-tables shows the results when playing one of our agent versions against the 3 finalist agents over the set of 200 games. Each entry shows the average score of each agent, and rows are ordered by ranking.

In both groups, adding the tariff market strategy and the wholesale market strategies resulted in significant improvements. Specifically, adding the tariff market strategy resulted in the largest improvements (agent U1). The next largest improvements were achieved when adding the wholesale market strategy (agent U9\_MDP). Adding the LWR-based prediction (U9\_MDP\_LWR) seems to be beneficial only for 2-broker games, possibly since its less conservative extrapolations work better with small number of competitors.

### 5.3 Chapter Summary

This chapter introduced TacTex-13, the champion power trading agent from the Power TAC 2013 finals. TacTex-13 uses an instantiation of LATTE that works effectively in practice. This instantiation is restricted to selling energy in the tariff market using fixed-rate tariffs, procuring energy solely in the wholesale market, and aiming for zero imbalance. These restrictions result in a computationally efficient implementation. TacTex-13 learns online to predict customer demand and wholesale costs and to bid in the wholesale market, using representations that allow it to learn with little data. In the 2013 finals, TacTex-13 won all of its 2-broker games and 15 out of the 16 4-broker games it completed successfully. Our experimental evaluation showed that two of TacTex-13's components were mainly responsible for its success: its tariff market strategy, induced by LATTE's lookahead, and its wholesale bidding policy. We have released the binary of TacTex-13, and it can be found at the following webpage: <http://www.cs.utexas.edu/users/TacTex/>. The next chapter introduces TacTex-15, which improves upon TacTex-13's demand prediction, cost prediction, and wholesale bidding policy.

## Chapter 6

# TacTex-15: A Winning Power Trading Agent

This chapter presents TacTex-15<sup>1</sup>, which is by many metrics the best Power TAC broker at the current time. Compared with TacTex-13, TacTex-15 presents three strategic improvements in its instantiation of LATTE, specifically in (a) demand prediction, (b) cost prediction, and (c) wholesale bidding policy. The first two improvements enhance the approximate transition and reward functions used by LATTE, and the third improvement enhances the implementation of the abstract wholesale action used in LATTE’s lookahead. The strategic improvements are relatively minor on the surface but result in large performance improvements. We start with describing the TacTex-15 agent (Section 6.1). Next, in the results section

---

<sup>0</sup>This chapter is based on a conference paper [112] (to appear) that I wrote with Professor Peter Stone. Author contributions were as follows: I was a Ph.D. student and did the complete implementation, and Peter was my advisor and collaborated with me on deciding on research directions and analyzing and interpreting results.

<sup>1</sup>TacTex-14 is not covered in this dissertation, since we consider it an exploratory work in progress towards TacTex-15. TacTex-14 extends TacTex-13 with different demand and cost predictors, specifically polynomial regression for customer subscription predictions and a non-parametric cost-predictor. TacTex-14 used early withdraw penalties, which were effective in the 2014 version of the simulator and became less effective in 2015. TacTex-14 is available to download from <http://www.cs.utexas.edu/users/TacTex/>

(Section 6.2) we report TacTex-15’s performance in competitions (Section 6.2.1) and in controlled experiments (Section 6.2.2), where using thousands of experiments we analyze the performance of TacTex-15 and the reasons for its success. TacTex-15’s binary and source code are publicly available (see Appendix A).

## 6.1 TacTex-15 Agent Description

TacTex-15 uses a restricted instantiation of LATTE that is similar to TacTex-13’s. Similarly to TacTex-13, TacTex-15 procures power in the wholesale market, sells power in the tariff market using fixed-rate tariffs, and aims for zero imbalance. Similarly to TacTex-13’s case, the reward components  $Q_t^{prod}$ ,  $p_t^{prod}$ ,  $I_{B_0,t}$ ,  $bal(I_{B_0,t})$ ,  $Q_t^{ask}$ , and  $p_t^{ask}$  are all zero, and  $Q_t^{bid} = Q_t^{wholesale} = Q_t^{cons}$ . Therefore TacTex-15’s instantiation of LATTE is quite similar to TacTex-13’s. On the other hand, TacTex-15 introduces three main improvements over TacTex-13, specifically in the instantiations of LATTE’s abstract functions (a) PredictTariffEffects, (b) PredictWholesalePrice, and (c) WholesaleBiddingPolicy. Improvements (a) and (b) result in a better transition and reward function models used by LATTE’s lookahead, and improvement (c) enhances the implementation of an abstract action used by LATTE’s lookahead, namely procuring a given energy quantity in the wholesale market. The function PredictTariffEffects is instantiated using a new *demand-predictor*, and the function PredictWholesalePrice is instantiated using a new *cost-predictor*, both are described next along with TacTex-15’s new wholesale bidding policy.

### 6.1.1 Instantiating PredictTariffEffects: a New Demand-Predictor

TacTex-15 instantiates LATTE’s PredictTariffEffects function using a new demand-predictor. This demand-predictor predicts customer subscription changes and fu-

ture demand, which determine  $Q_t^{cons}$ ,  $p_t^{cons}$ , and  $Q_t^{\oplus cons}$  in line 7 of Algorithm 1.<sup>2</sup> TacTex-13 learned a demand-predictor from data. In general, learning a demand-predictor helps in adapting to new or changing environments, but in Power TAC there is no need to do so: these complex stochastic customer behaviors are coded in Power TAC’s open-source simulator. Instead, TacTex-15 uses the simulator’s customer code as a basis for its demand-predictor. However, this code does not provide a complete demand-predictor: it relies on information hidden from brokers. TacTex-15 heuristically seeds this information to reasonable values: customers of other brokers are assumed to be subscribed to the best tariffs, customer subscriptions changes are predicted in the limit (expected values after infinite time, thus ignoring customer inertia and limited-time tariff-binding fees), and customer demand parameters are set to expected values. Using the customers’ code as a basis for TacTex-15’s demand-predictor allows us to examine the importance of having an accurate demand-prediction to the LATTE’s overall performance (see Section 6.2.2).

### 6.1.2 Instantiating PredictWholesalePrice: a New Cost-Predictor

As we have seen at the beginning of this chapter, in TacTex-15’s implementation of LATTE  $Q_t^{bid} = Q_t^{wholesale} = Q_t^{cons}$  (to remind,  $Q_t^{bid}$  is the quantity procured in the wholesale market,  $Q_t^{wholesale}$  is the net quantity traded in the wholesale market, and  $Q_t^{cons}$  is the quantity consumed by TacTex-15’s customers; here we overload notation and treat these variables as actual values when  $t$  is a past timeslot and as predicted values when  $t$  is a future timeslot). This means that TacTex-15 predicts its abstract wholesale actions to be procurement actions. Therefore, TacTex-15 implements  $\text{PredictWholesalePrice}(Q_t^{wholesale}, Q_t^{\oplus wholesale}, t)$  using a wholesale *cost-predictor*, which is described in this section.

TacTex-15’s cost-predictor needs to be learned, since costs are determined by

---

<sup>2</sup>TacTex-13 did not predict  $Q_t^{\oplus cons}$  since it did not need it for its instantiation of LATTE. TacTex-15 needs this information for predicting wholesale costs (see below).

brokers' bidding strategies and traded quantities, which are unknown in advance and may change dynamically. Therefore, TacTex-15 learns and adapts a cost-predictor online. A cost predictor ultimately needs to predict the price  $p_t^{bid}$  of procuring a quantity  $Q_t^{bid}$  in the wholesale market. In a monopoly setup, where a single broker bids to procure energy from generation companies,  $Q_t^{bid}$  is predictive of  $p_t^{bid}$ . However, in a competitive environment with multiple brokers, using  $Q_t^{bid}$  by itself does not provide enough information to reliably predict  $p_t^{bid}$ :  $p_t^{bid}$  is determined by additional features such as the total quantities traded by other brokers; how each of these quantities is divided between the 24 auctions in which energy is traded for the target timeslot  $t$ ; which brokers participate in each of the 24 auctions (since each broker has its own bidding policy); whether some brokers resell energy they had procured in earlier auctions due to customers migrating to their competitors; and how customer subscriptions are divided among brokers (since it determines the quantities brokers need to trade). This information is encapsulated in the state variables  $\mathcal{S}$  and  $\mathcal{A}$ , and is hidden from TacTex-15.

To bypass the problem of predicting  $p_t^{bid}$  from unobserved features, TacTex-15 predicts  $p_t^{bid}$  from observable information that is assumed to be correlated with unobserved information that causally affect these unobserved features and therefore  $p_t^{bid}$ . Specifically, TacTex-15 assumes that given brokers' bidding strategies: (i) most of these unobserved features are causally determined by brokers' predictions of their customers' consumption and production (one exception is customer subscriptions, which likely *determine* brokers' predictions), (ii) brokers' predictions are correlated with the actual consumption and production of their customers, and (iii) the actual consumption and production of competitors' customers are correlated with TacTex-15's prediction of them. Therefore, TacTex-15's cost-predictor predicts  $p_t^{bid}$  from TacTex-15's predictions of competitors' production and consumption and of its own customers' consumption.

TacTex-15 needs to learn such a cost-predictor from past data of actual consumption, production and prices, but individual past consumption and production of competitors' customers are unobserved by TacTex-15. However, their past *total* consumption and *total* production are observed by TacTex-15: every broker receives a report on the total consumption and production in each timeslot, so TacTex-15 subtracts its customers' consumption and production from these totals to figure out competitors' totals. Therefore, TacTex-15 learns to predict wholesale costs from its own and from competitors' past total consumption and production. Specifically, TacTex-15 learns a cost-predictor from past data of the form  $\langle Q_t^{cons}, Q_t^{\oplus cons} - Q_t^{\oplus prod} \rangle \rightarrow p_t^{bid}$  (which is  $\langle Q_t^{cons} - Q_t^{prod}, Q_t^{\oplus cons} - Q_t^{\oplus prod} \rangle \rightarrow p_t^{bid}$  with  $Q_t^{prod} = 0$ ). To use this predictor to predict  $p_t^{bid}$  for a future timeslot  $t$ , TacTex-15 needs to provide the features  $\langle Q_t^{cons}, Q_t^{\oplus cons} - Q_t^{\oplus prod} \rangle$  for this future timeslot; it does so using the predictions coming from the demand predictor. We note that using informal experimentation, we confirmed that while  $Q_t^{bid}$  (and the corresponding  $Q_t^{cons}$ ) is uncorrelated with  $p_t^{bid}$ , the combination  $\langle Q_t^{cons}, Q_t^{\oplus cons} - Q_t^{\oplus prod} \rangle$  is correlated with  $p_t^{bid}$ .

TacTex-15's instantiation of  $\text{PredictWholesalePrice}(Q_t^{wholesale}, Q_t^{\oplus wholesale}, t)$  is described in Algorithm 5. Recall that in TacTex-15's instantiation of LATTE,  $Q_t^{wholesale} = Q_t^{cons}$  and  $Q_t^{\oplus wholesale} = Q_t^{\oplus cons} - Q_t^{\oplus prod}$ . Therefore the function  $\text{PredictWholesalePrice}$  receives the features  $\langle Q_t^{cons}, Q_t^{\oplus cons} - Q_t^{\oplus prod} \rangle$  as parameters. TacTex-15 implements  $\text{PredictWholesalePrice}$  using an adaptive cost-predictor, which has two components: a linear regression predictor trained on *boot data* (created once in line 2 and used in line 6), and a real-time correction factor constructed from the last 24 hours' prediction errors (line 7). The boot data is sent by the simulator at game start, and contains wholesale transactions made by a default simulated broker as a single buyer in the market. This boot data serves as a basis for a 1-dimensional regression that maps wholesale quantities to prices (trained in line



2). To be able to use this 1-dimensional regression, we construct a 1-dimensional feature from  $Q_t^{wholesale}, Q_t^{\oplus wholesale}$  by summing them. Since the correction factor is constructed from little data (to ensure responsiveness), we limit it to bias correction. The boot data is larger (336 instances) so we use it to determine the slope. TacTex-13's cost-predictor ignored traded quantities, and predicted past average prices based on time. We compare the two predictors in Section 6.2.2.

---

**Algorithm 5** TacTex-15::PredictWholesalePrice( $Q_t^{wholesale}, Q_t^{\oplus wholesale}, t$ )

---

```

1: if firstTime then
2:   cost-predictor.trainLinearRegression( $\{\langle Q_{t'}^{bid}, p_{t'}^{bid} \rangle\}_{t' \in bootdata}$ )
3:   firstTime  $\leftarrow$  false
4: return cost-predictor.predict( $Q_t^{wholesale} + Q_t^{\oplus wholesale}$ )

```

---

▷ The cost-predictor function used by PredictWholesalePrice

---

```

5: function cost-predictor.predict( $Q$ )
6:   price  $\leftarrow$  cost-predictor.getLinearRegression.predict( $Q$ )
7:   correctionFactor  $\leftarrow$  cost-predictor.averagePredictionErrorInLast24Hours()
8:   return price - correctionFactor

```

---

### 6.1.3 Instantiating WholesaleBiddingPolicy

In the wholesale market, TacTex-15 hedges between truthful and strategic (i.e. non-truthful) bidding. TacTex-15's value of procured energy unit in the wholesale market is the imbalance fee, which is the price TacTex-15 would pay if it takes no procurement action. Therefore, a truthful bid by TacTex-15 would have a limit price of the predicted imbalance fee, denoted here as  $\bar{p}$ . By bidding truthfully, TacTex-15 would get the highest priority among competitors who bid less than  $\bar{p}$  and would never pay more than  $\bar{p}$ . However, since the sequential double-auction mechanism is not incentive compatible, truthful bidding is suboptimal in some situations. For instance, when the truthful bid is setting the clearing price (i.e. when it is lowest-priced cleared bid), bidding lower could be better (as long as it is above the highest-cleared ask), since it reduces the clearing price.

TacTex-13 used an optimistic strategic (i.e. non-truthful) sequential bidding policy  $\pi(Q)$ : it assumed that any bid with limit price higher than the clearing price would result in procurement of the complete quantity specified in the bid. Therefore, this policy set its limit prices to be slightly higher than the highest clearing price it was willing to pay in each of the 24 MDP states. This policy is optimal in some situations (e.g. single-buyer or cooperative setups), but can be exploited by competitors who learn to bid slightly higher: such competitors could benefit from the low prices this policy sets, while getting higher priority in the auction and therefore getting a larger fraction of cheap energy than the user of this policy, whose bids may end up being only partially cleared.

Since each of the two strategies is beneficial in different situations, TacTex-15 hedges between them. Let  $\underline{p}$  be the limit price suggested by TacTex-13's policy, and  $\epsilon$  be the minimum amount that can be traded (0.01 MWh in Power TAC). To bid for a quantity  $Q_t^{bid}$ , TacTex-15 submits the following 25 orders (each order is of the form  $\langle energyAmount, limitPrice, targetTime \rangle$ ):  $\langle Q_t^{bid} - 24\epsilon, \bar{p}, t \rangle, \{ \langle \epsilon, \underline{p} + i \frac{\bar{p} - \underline{p}}{24}, t \rangle \}_{i=0}^{23}$ . This policy benefits from both worlds: if TacTex-15 sets the price, it will either be the strategic price  $\underline{p}$  returned by  $\pi(Q)$ , or the lowest among its higher bids. If another broker sets the price, TacTex-15 will have a higher priority and benefit from the lower price as long as it is not higher than  $\bar{p}$ .

## 6.2 Results

We analyze TacTex-15's performance in competitions (Section 6.2.1) and controlled experiments (Section 6.2.2).

### 6.2.1 Competition Results: Power TAC 2015 Finals Analysis

The Power TAC 2015 Finals included 11 teams from universities in America, Europe and Asia. 230 games were played continually over a week, in three different

sizes: 3-brokers, 9-brokers, and 11-brokers. A day after the finals ended, 8 of the teams competed in a post-finals, demo-competition with 70 4-broker games. While being unofficial, this competition was run similarly to the finals with one important difference: a simulator-loophole that was exploited during the finals, was fixed. Due to the proximity to the finals, and a parallel workshop, we believe that teams used the same brokers they used in the finals.

Table 6.1 summarizes the 2015 finals results. While **TacTex-15** was officially ranked 2nd, it was the best broker that did not exploit a simulator-loophole: the 1st-ranked broker gained the highest overall score by exploiting a simulator loophole in 3-broker games, which resulted in unrealistic dynamics and an unrealistically high score that biased the final ranking (see dark gray cells in Table 6.1).<sup>3</sup> Specifically, Maxon15 subscribed customers to inflated tariffs which promised customers large payments if customers *unsubscribed* from them after a period shorter than a single timeslot. However, customers had no way to unsubscribe quickly enough (specifically after less than a single timeslot) to collect the promised payments. Nevertheless, due to the loophole, customers subscribed to these tariffs assuming they could collect the payments, and paid inflated prices to Maxon15.

After the finals, the loophole was fixed. When replaying 3-broker competition games without the loophole, Maxon15 no longer won by a large gap, but instead lost by a large gap to **TacTex-15**. When taking into account only 11- and 9-broker games from the finals (where the loophole had no impact), **TacTex-15** ended 1st with a total z-score of 0.142 ahead of CUHKTac and 0.551 ahead of Maxon15, finishing slightly behind CUHKTac in 11-broker games (by 0.065) and ahead of CUHKTac in 9-broker games (by 0.207). In the post-finals demo competition with a repaired simulator, **TacTex-15** won by a large gap ahead of the others (Table 6.2), making 50% more profits than the 2nd place (Maxon15). Maxon15 used the same strategy

---

<sup>3</sup>Maxon was not disqualified: they explained it as an unintended result of automatic parameter tuning right before the finals.

Table 6.1: **Power TAC 2015 finals results.** Ranking is determined by the “Total” score, which is a sum of individual z-scores in each game size, displayed in the columns “11-brokers” (10 games played by all brokers), “9-brokers” (45 games played by each broker) and “3-brokers” (45 games played by each broker).

Broker	11-brokers	9-brokers	3-brokers	Total
Maxon15	0.611	0.801	1.990	3.402
TacTex-15	0.897	1.066	0.258	2.221
CUHKTac	0.962	0.859	0.106	1.927
AgentUDE	0.421	0.367	0.809	1.597
Sharpy	0.429	0.614	0.521	1.564
COLDPower	0.726	0.397	-0.751	0.371
cwiBroker	-0.002	-0.120	0.465	0.343
Mertacor	0.413	0.142	-1.341	-0.786
NTUTacAgent	-1.017	-1.638	0.453	-2.202
SPOT	-1.052	-0.243	-1.032	-2.327
CrocodileAgent	-2.387	-2.244	-1.479	-6.111

Table 6.2: **Power TAC 2015 post-finals demo competition results.** 70 games were played in a single game-size (4-brokers). Ranking is determined by z-score.

Broker	4-brokers (profits)	4-brokers (z-score)
TacTex-15	15.0M	1.122
Maxon15	10.7M	0.627
CUHKTac	10.0M	0.537
AgentUDE	9.7M	0.509
cwiBroker2015	7.9M	0.297
Sharpy	4.6M	-0.092
COLDPower	-0.8M	-0.724
SPOT	-14.0M	-2.276

as before, but it was not as effective with the loophole fixed.<sup>4</sup>

Figure 6.1 shows an analysis of TacTex-15's performance in the 2015 finals and in the post-finals competition. In 11-broker games CUHKTac (1st) and TacTex-15 (2nd) won by a large gap over the other brokers, where most brokers ended with losses. In 9-broker games TacTex-15 won by a large gap, making 30% more profit than the 2nd place broker in this category (CUHKTac), despite missing 3 out of 45 games due to network connection problems. The revenue and costs plots show that in 11- and 9-broker games TacTex-15 chose to reduce its market share, likely due to the fierce competition, so that its revenue and costs were lower compared with other top brokers, while its profit remained high. In 3-broker games TacTex-15 typically performed the best, although this is harder to see in the figure, due to several events that biased the final averages: (a) Maxon15's loophole-exploitation, discussed above; (b) About 1/2 of AgentUDE's, Sharpy's and cwiBroker's 3-broker game scores come from single outlier games in which they played against a non-functioning broker (Mertacor) and/or a competitor's crash in a monopoly/duopoly situation; (c) while NTUTacAgent did not exploit the simulator's loophole, the loophole indirectly inflated its final score (d) TacTex-15 missed 5 out of its 45 3-broker games due to network connection problems, resulting in a score of 0 in these games, and a reduction of 4.3% in TacTex-15's average profit. In the 4-broker games of the post-finals competition TacTex-15 made about 50% more profit than the 2nd place broker. The revenue and costs plots show that it had a similar revenue to the 2nd and 3rd place brokers, but much lower costs; higher revenue and lower costs than the 5th, 6th brokers; and almost double the revenue of each of the other brokers.

---

<sup>4</sup>To be fair, one should note that they did not retune their parameters to the repaired simulator. On the other hand, it's not clear that other parameters would have done particularly better in the absence of the loophole.

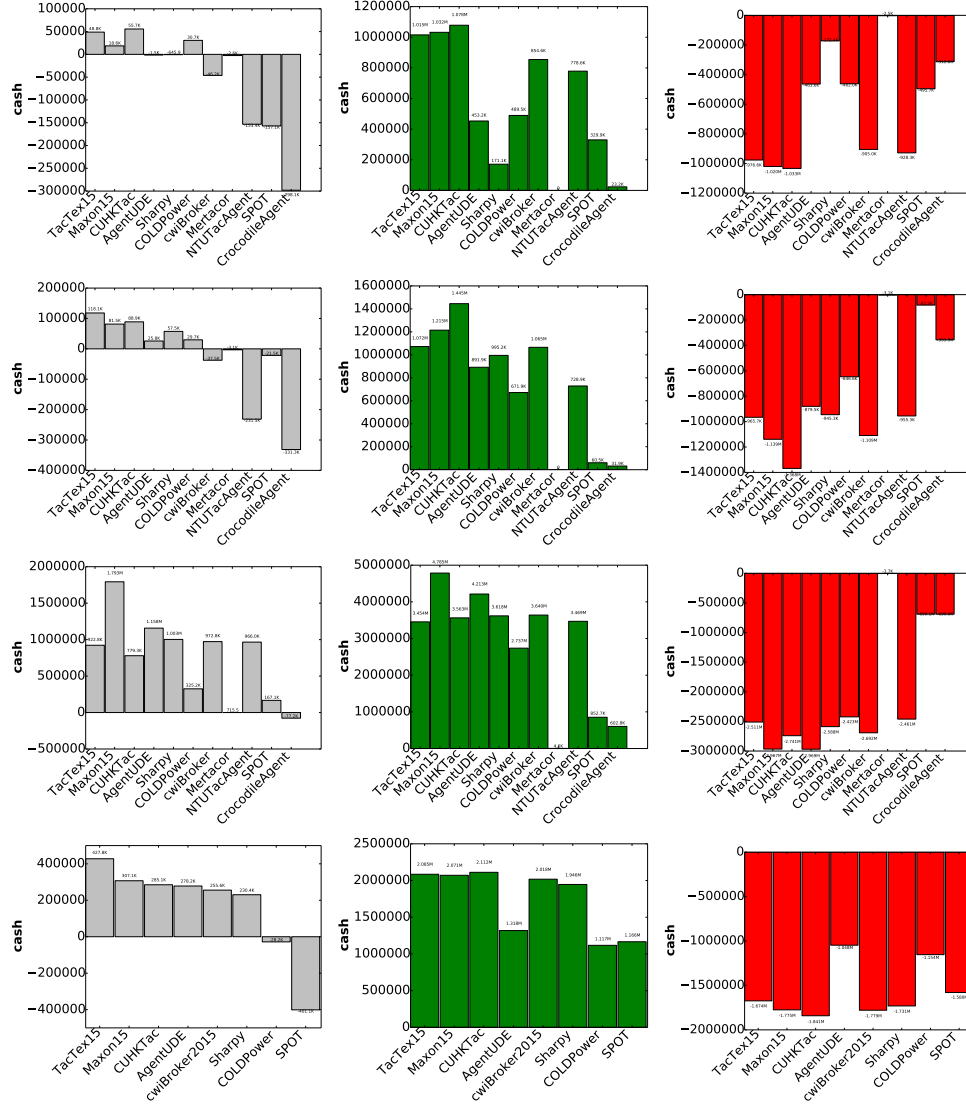


Figure 6.1: **2015 competition analysis: average profit, revenue and costs.** The top 3 lines respectively summarize 11-, 9-, 3-broker games from Power TAC 2015 finals; the bottom line summarizes the 4-broker games of the post-finals demo competition. Each line shows average profit (left), revenue (middle), and costs (right). 3-broker game results are biased due to (a) a simulator-loop-hole exploitation by Maxon15; (b) an outlier game for AgentUDE, cwiBroker and Sharpy; (c) a simulator-bug (NTUTacAgent); and (d) TacTex-15 missing 5 games due to network problems. More details are in the text.

### 6.2.2 Controlled Experiments

While the competition is motivating and its results are illustrative, it cannot isolate specific broker components in a statistically significant way. We therefore subsequently tested TacTex-15 in thousands of games, in two types of controlled experiments: (a) performance tests, and (b) ablation analysis tests, which evaluate the contribution of TacTex-15’s main components to its overall performance. The simulator version used in the experiments in this section is specified in Appendix A.

#### Experimental Setup

Each experiment consisted of running 56 games against a set of opponent brokers, using broker binaries of 2015 finalists. To better evaluate statistical significance, we held most of the random factors in the simulation fixed across experiments (random seeds, weather conditions). To fix weather conditions, we used weather files containing 3 months of real-world weather. To cover year-round weather conditions we used 8 weather files (each file used by 1/8 of the games) with start-dates of January, April, July, October of 2009 and 2010.

#### Performance Tests

A successful broker should perform well in expectation against every set of opponents, under different stochastic conditions (here weather/random seeds). At the time of running the experiments, five 2015 finalists had released their brokers’ binaries. We used these binaries to test TacTex-15’s performance in 2, 3, 4, 5, 6-broker games. We generated combinations of brokers for each game size, and tested each combination in 56 games, as described above. Figure 6.2 presents the results. TacTex-15 significantly won against every combination of opponents, typically by a large gap.

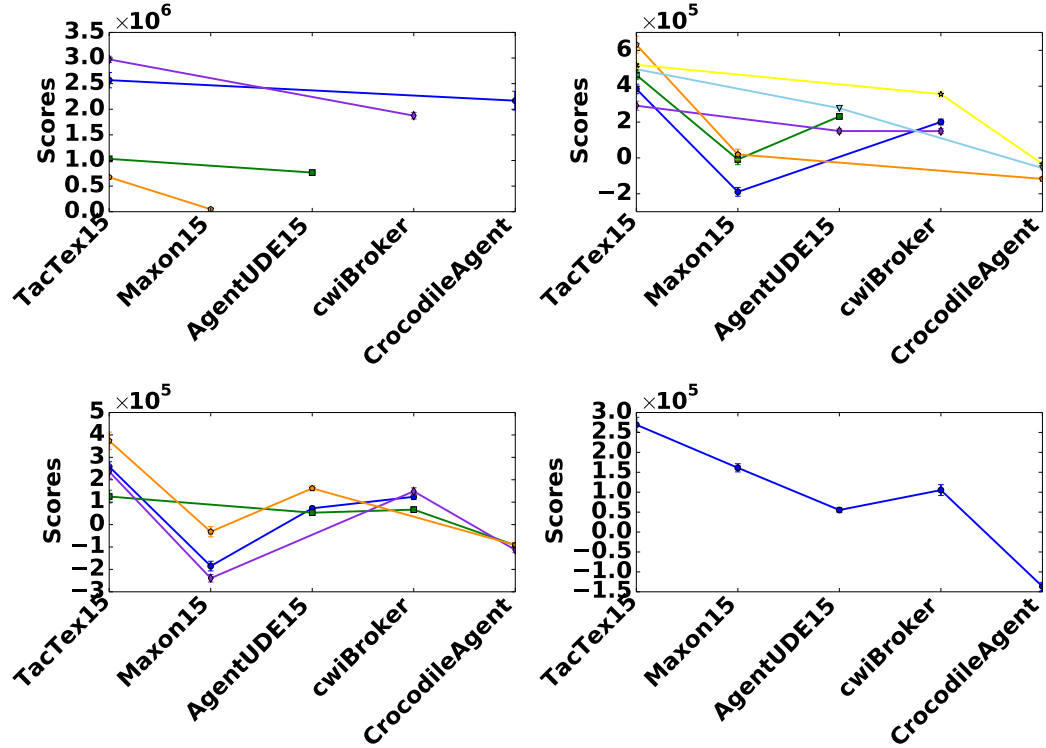


Figure 6.2: **Performance of TacTex-15 against Power TAC 2015 finalists in controlled experiments of game-sizes of 2-5.** Each line represents the average scores of a combination of brokers playing each other under a variety of conditions (note the small error bars). Results are shown for game-sizes of 2-, 3-, 4-, 5-brokers (top-left, top-right, bottom-left, bottom-right, respectively). Similar results for 6-brokers are omitted. TacTex-15 consistently won against all combinations of brokers, in all game-sizes.



## Ablation Analysis

To understand the reasons for TacTex-15’s success, we tested the contribution of TacTex-15’s main components to its overall performance, in all possible game-sizes (2,...,6). We created three ablated versions of TacTex-15 by disabling each of its main components. For each game size, we selected the “strongest” combination of opponents, against which TacTex-15 had the lowest score. We tested each ablated version against these opponents in a 56-game experiment, holding random seeds and weather conditions fixed to the same values used against TacTex-15. When disabling a component, we used as a baseline the corresponding component used by TacTex-13 (since TacTex-15’s ablated version must have some component in place of a disabled one to run properly). Figure 6.3 shows the results of our ablation analysis. Disabling the cost-predictor (Abl-cost) did not have significant impact on TacTex-15’s performance (however it can reduce performance, see Figure 6.8). Disabling the wholesale-bidding policy (Abl-bid) significantly hurts TacTex-15’s performance: it reduces TacTex-15’s score in game sizes 2, 4, 5, 6, and it causes TacTex-15 to either lose its lead (in game sizes 2, 3) or have a smaller victory margin (in game sizes 4, 5, 6). Disabling the demand-predictor (Abl-demand) significantly hurts TacTex-15’s performance: it drops TacTex-15’s score in all game sizes, and causes TacTex-15 to either lose its lead (in game sizes 3, 5, 6) or have a smaller victory margin (in game sizes 2, 4).

## Ablation Analysis Extensions

To gain more insight into the importance of TacTex-15’s main components, we extended each ablation experiment. First, we extended TacTex-15’s demand-predictor ablation analysis from a binary ablation test (disabled/enabled, see Abl-demand in Figure 6.3) to a continuum of ablation-levels, thus testing TacTex-15’s sensitivity to demand prediction errors. Figure 6.4 shows the performance-degradation as a

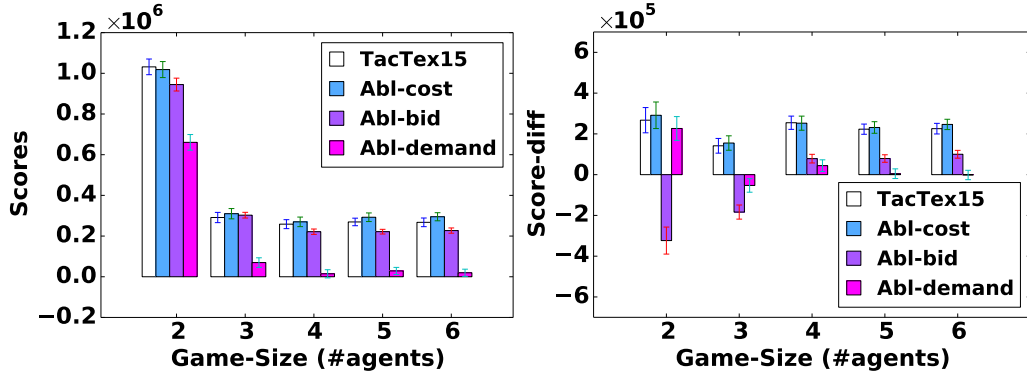


Figure 6.3: **Ablation analysis for 2-6 broker games.** The performance of TacTex-15 is compared with three of its ablated versions, when playing against the strongest combination of opponents in each game size. Ablated versions are constructed from TacTex-15 by disabling cost predictor (Abl-cost), wholesale-bidding policy (Abl-bid), and demand-predictor (Abl-demand). The left figure shows the average scores of each version in each game size; the right figure shows the average score-differences of each version from opponents’ average score (y-axes’ scales are the same).

function of ablation-level. We see that TacTex-15’s degrades quickly even for small levels of ablation. We conclude that having an accurate demand-predictor is crucial for the success of TacTex-15’s implementation of LATTE.

To better understand why the broker’s performance degraded when ablating the demand-predictor, we examined more closely the games played by TacTex-15 (left-most point in Figure 6.4) and by Abl-demand (right-most point in Figure 6.4). Figure 6.5 shows the revenue and cost components that compose the brokers’ score (cash). In this figure, we see that: (a) Abl-demand pays more tariff publication fees than TacTex-15 (by about 148K), (b) Abl-demand distributes (sells) more energy than TacTex-15, thus paying more distribution fees than TacTex-15 (by about 33K), (c) energy selling income (from consumption tariffs, denoted ConsTariff) is similar between TacTex-15 and Abl-demand, (d) energy procurement costs (wholesale and balancing) are similar between TacTex-15 and Abl-demand. In addition, TacTex-15’s average energy selling price was 4.3% higher than that of Abl-demand (7.45 vs. 7.14

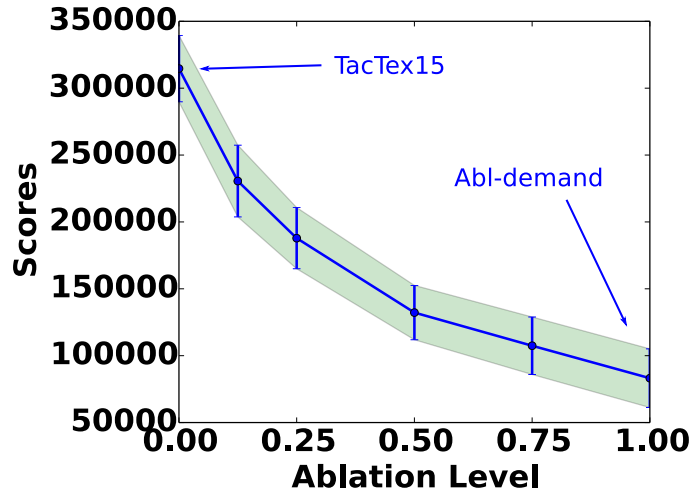


Figure 6.4: **Performance as a function of ablation level of the demand-predictor in 3-agent games.** The plot shows the degradation in TacTex-15’s performance as the ablation level of its demand predictor increases. To change ablation level along a continuum, TacTex-15 uses a weighted combination of two demand-predictors: (1) its own predictor, and (2) TacTex-13’s demand-predictor, which was used by the ablated agent Abl-demand in Figure 6.3. Ablation level is then represented as the relative weight given to predictor (2), so that a weight of 0 means “no-ablation”, and a weight of 1 means “full-ablation”.

cents/kWh, not seen in the figure). We conclude that ablating the demand-predictor (and using TacTex-13’s demand-predictor instead) causes the broker to publish too many tariffs, at too low prices. This conclusion is confirmed by examining the broker’s actions in a single game: Figure 6.6 shows how TacTex-15 stops publishing tariffs early in the game, while Abl-demand keeps publishing tariffs and reducing prices until the end of the game. Figure 6.7 shows that Abl-demand typically has a larger number of customers than TacTex-15, however its profit is lower and flattens toward the end of the game, while TacTex-15’s profit keeps increasing. We conclude that ablating TacTex-15’s demand-predictor results in over-estimation of the utility predicted for tariff price reductions, compared with the utility predicted for taking no action.

Next, we extended the ablation analysis of TacTex-15’s wholesale-bidding policy with additional comparisons against its ablated version (used by Abl-bid, see Figure 6.3). Abl-bid’s policy (which is TacTex-13’s policy) can be viewed as more cooperative than TacTex-15’s, since it submits lower bids, and thus may result in lower costs against an opponent using a similar policy. To understand whether Abl-bid’s cooperative policy is preferable in some situations, we created a payoff matrix (Table 6.3) by running 2-broker games, testing both TacTex-15 and Abl-bid in self-play and against each other. While Abl-bid’s cooperative policy indeed resulted in lower costs in self-play (40 \$/MWh vs. 57 \$/MWh, a 29.8% reduction), Abl-bid’s total scores in self-play were not higher than TacTex-15’s, since the competitive selling policy reduced selling-prices further than TacTex-15’s, such that the profit remained similar to TacTex-15’s. As a result, TacTex-15’s competitive policy dominated Abl-bid’s cooperative policy in Table 6.3’s experiments.

We ran additional self-play experiments using 3-, 4-, 5-broker games. In these cases Abl-bid’s more cooperative bidding policy resulted in higher scores than TacTex-15, mainly since Abl-bid’s lower energy costs enabled a longer price-reduction

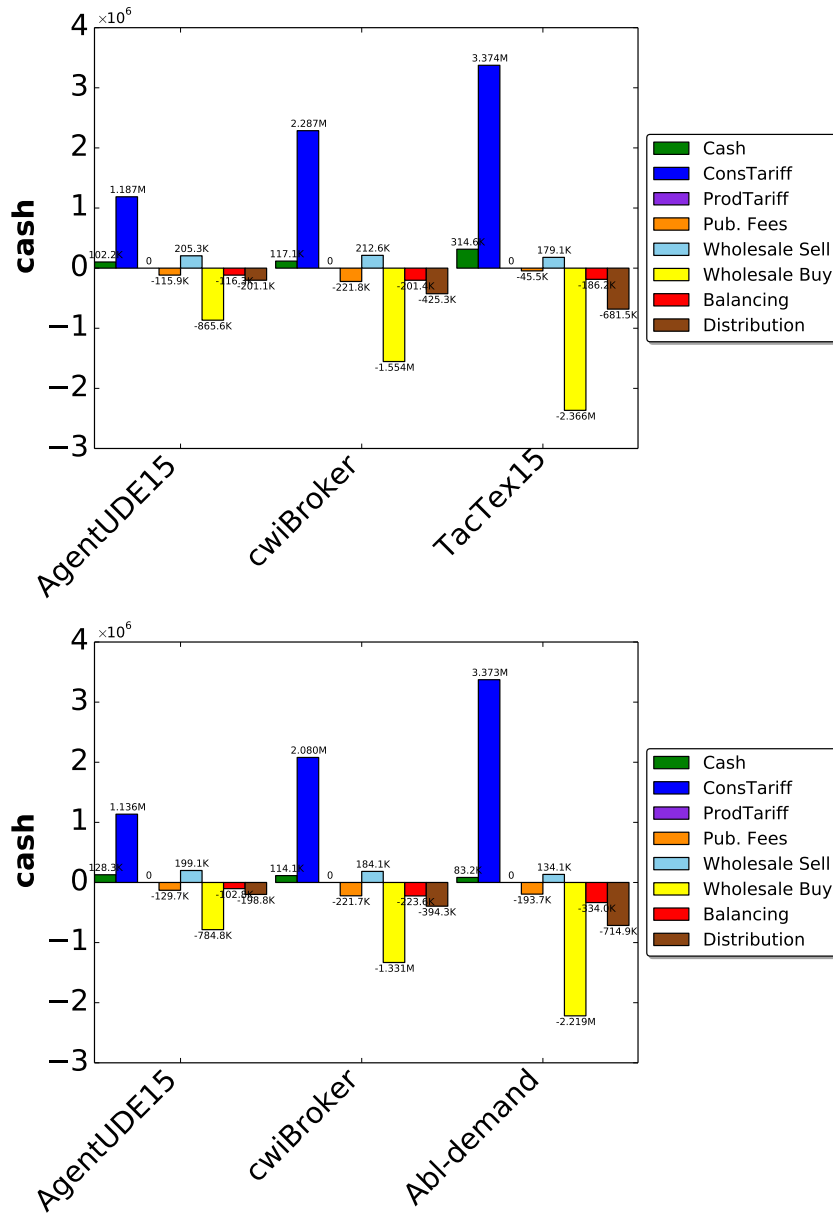


Figure 6.5: **Demand-predictor ablation: revenue and cost components.** The bar charts show the average revenue and cost components when each of TacTex-15 (top) and its demand-predictor ablated version, (Abl-demand, bottom), played 3-broker games against AgentUDE15 and cwiBroker15. The top bar chart corresponds to the left-most point in Figure 6.4 and the bottom bar chart corresponds to the right-most point in Figure 6.4.

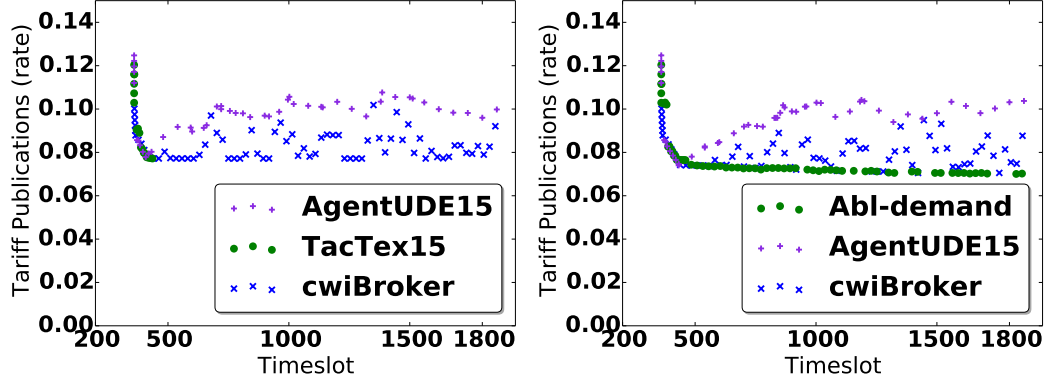


Figure 6.6: **Demand-predictor ablation: broker behavior in an example game.** The plots show tariff publications throughout the game, in two repetitions of a game using the same opponents, random seeds, and weather conditions, first with TacTex-15 (top) and then with its demand-predictor ablated version (Abl-demand, bottom). A tariff action is represented by the timeslot in which it took place (x-axis), and by the tariff’s selling price (y-axis). A game normally starts at timeslot 360 (since timeslots 0-359 are “bootstrap” period).

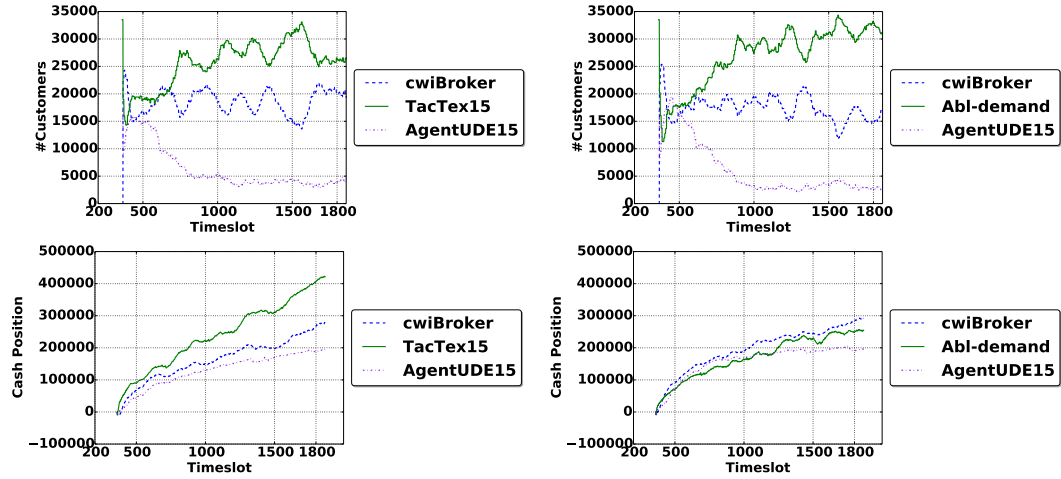


Figure 6.7: **Demand-predictor ablation: customers and profits in an example game.** The plot shows customer subscriptions (top row) and profits (cash, bottom row) throughout a game, in the game-repetitions described in Figure 6.6. The plots from the game played with TacTex-15 (top plot in Figure 6.6) are in the left column, and the plots from the game played with Abl-demand (bottom plot in Figure 6.6) are in the right column.

period after game-start, during which selling-prices were higher than the eventual equilibrium after which the profit of all brokers increased in the same pace.

**Table 6.3: Payoff matrix of two wholesale-bidding strategies in 2-agent games.** The matrix shows a game-theoretic payoff matrix of two wholesale bidding strategies: (a) **Comp-Bid** is TacTex-15’s competitive bidding policy, and (b) **Coop-Bid** is Abl-bid’s (and TacTex-13’s) cooperative bidding policy from Figure 6.3. The matrix entries show the average scores of agents using these strategies (TacTex-15 and Abl-bid, respectively) in self-play and against each other.

Payoff Matrix		
	Coop-Bid	Comp-Bid
Coop-Bid	1.0M 1.0M	1.6M 0.8M
Comp-Bid	0.8M 1.6M	1.0M 1.0M

Finally, we extended TacTex-15’s cost-predictor ablation analysis. Even though ablating TacTex-15’s cost predictor did not reduce performance against the 2015 finalists (Figure 6.3), we expect it to reduce performance when wholesale costs change more dynamically. Figure 6.8 shows the result of such an experiment, where TacTex-15 played against its cost-predictor ablated version (Abl-cost from Figure 6.3), and was quicker to react to a drop in wholesale costs and thus significantly won against Abl-cost.

## 6.3 Chapter Summary

This chapter introduced TacTex-15, which extends TacTex-13’s implementation of LATTE. TacTex-15 consistently achieved top performance in Power TAC 2015 competitions and in extensive controlled experiments. Specifically, (i) in the Power TAC 2015 finals it was the best agent that did not exploit a simulator loophole, (ii) in the post-finals demo competition it won by a large gap over the rest of the agents, making 50% more profit than the second place agent, and (iii) in controlled experiments

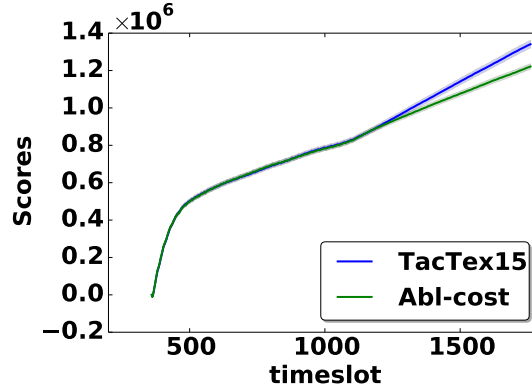


Figure 6.8: **Cost-predictor ablation in presence of abruptly changing market-costs.** The plot shows the average cumulative profit (with confidence bounds) as a function of time in head-to-head games of TacTex-15 vs. its cost-predictor ablated version (Abl-cost from Figure 6.3), when market costs abruptly dropped in timeslot 1080 (mid-game). TacTex-15 was quicker to react due to its more adaptive cost-predictor: it reduced selling prices, and thus gained market-share and increased its profits. To create a market-cost drop effect, we could reduce either the sellers’ asks, or the brokers’ bids. We implemented the latter (to avoid changing the simulator), by making both brokers switch their bidding policies in timeslot 1080 from competitive policies (of TacTex-15) to a cooperative policy (of TacTex-13).

it significantly won against every combination of competitors. TacTex-15 improves upon TacTex-13’s instantiation of LATTE in three main ways, where two of them improve the predictive model used in LATTE’s lookahead (its demand-predictor and cost-predictor) and one of them improves an implementation of an abstract action used by LATTE (its wholesale bidding policy). These improvements turn out to be critical to TACTEX’s performance. Specifically, disabling its demand-predictor and wholesale bidding policy reduced its performance in 2,3,4,5-broker games, and disabling its cost-predictor reduced performance in games in which wholesale costs changed abruptly. The next chapter describes an extension of TacTex-15’s implementation of LATTE, which searches over a superset of tariff actions that includes *Time-Of-Use* tariffs, which are a proposed method for implementing *demand-response* in



future electricity markets (i.e. affecting customer demand to adapt to supply conditions).

## Chapter 7

# Autonomous Electricity Trading using Time-Of-Use Tariffs in a Competitive Market

One of the milestones in the smart grid vision is “customer participation in power markets through demand-side-management” [114]. Demand-side management (DSM) refers to adapting customer demand to supply conditions. One of the main methods proposed for implementing DSM is *Time-Of-Use (TOU) pricing* [34, 49], which specifies time-of-day-based electricity prices (e.g. hourly prices), in contrast to the fixed prices that currently dominate retail electricity markets. As of the current date, TOU pricing schemes are proposed to take effect in California starting January 2019 [91]. TOU pricing incentivizes customers to adapt their consumption and shift portions of it to cheaper times. Shifting consumption can reduce customers’ costs while potentially increasing their discomfort.

---

<sup>0</sup>This chapter is based on a published conference paper [113] that I wrote with Professor Peter Stone. Author contributions were as follows: I was a Ph.D. student and did the complete implementation, and Peter was my advisor and collaborated with me on deciding on research directions and analyzing and interpreting results.

In Power TAC, fixed pricing is implemented through fixed-rate tariffs, and TOU pricing is implemented through TOU tariffs. Chapters 5 and 6 described two successful instantiations of LATTE used by TacTex-13 and TacTex-15, which achieved top performance in competitions and controlled experiments. These instantiations sold electricity to retail consumers only through fixed-rate tariffs, in large part because the Power TAC simulator did not have demand-shifting customers until early 2015. This chapter describes a third instantiation of LATTE, called LATTE-TOU, which extends these previous instantiations of LATTE by allowing the broker to use TOU tariffs in the tariff market. The broker binaries and simulator versions used in this chapter are publicly available, along with source code that can be configured through text files to run LATTE-TOU and behave like these released binaries (see Appendix A). This chapter’s primary contributions are:

- We enhance the Power TAC simulator’s consumers with demand-shifting capabilities, and enable demand-shifting for about 50,000 simulated consumers. Specifically, we enhance Power TAC’s factored-customer models (Section 2.2.2) (a) to evaluate a TOU tariff’s utility based on the cost and discomfort of the predicted shifted-consumption under this TOU tariff, and (b) to allow for different consumption profiles for different members of the population, each based on the tariff the member is subscribed to.
- We extend LATTE to reason effectively about TOU Tariffs.
- We show that the problem of optimizing TOU tariffs in competitive markets is intractable, and propose an efficient optimization algorithm that approximates its solution. Our algorithm is fully implemented in a new instantiation of LATTE, called LATTE-TOU, which is used by our broker agent.
- LATTE-TOU leads to 15% peak-demand reduction. To the best of our knowledge, our work is the first to show that TOU can achieve the primary goal

of peak-flattening in competitive markets in a large-scale, realistic simulation such as Power TAC. Our agent’s peak-flattening results in greater profits and/or profit-share and allows it to beat fixed-rate brokers, specifically the 1st and 2nd place agents from the 2014 Power TAC finals, while reducing the electricity costs of both its customers and its competitors’ customers.

- Using extensive experimentation, we analyze several economic implications of using TOU in competitive retail markets. For instance, while previous research warned that TOU tariffs could induce *customer-herding*, our TOU broker prevented it by implicitly coordinating flattening through profit-maximizing tariffs. Our broker’s prevention of customer herding underlines a potential benefit of employing autonomous TOU brokers in competitive power markets.

## 7.1 Background

This section elaborates on the motivation for DSM and TOU tariffs (Section 7.1.1), and describes how Power TAC customers react to DSM using TOU tariffs (Section 7.1.2).

### 7.1.1 Motivation: DSM and TOU Tariffs

A main motivator for DSM in general, and for TOU tariffs in particular, is the variability in electricity generation prices. Electricity generation prices depend on the types of electricity generators used, which frequently depend on energy availability and predicted daily demand patterns. Figure 7.1 shows the marginal electricity generation costs as a function of generated power and generation type in three large US wholesale markets: ERCOT, PJM, and CAISO. Since these marginal electricity generation costs are the wholesale offer prices (asks) of generators, Figure 7.1 shows the costs of procuring electricity in the wholesale market. The figure demon-

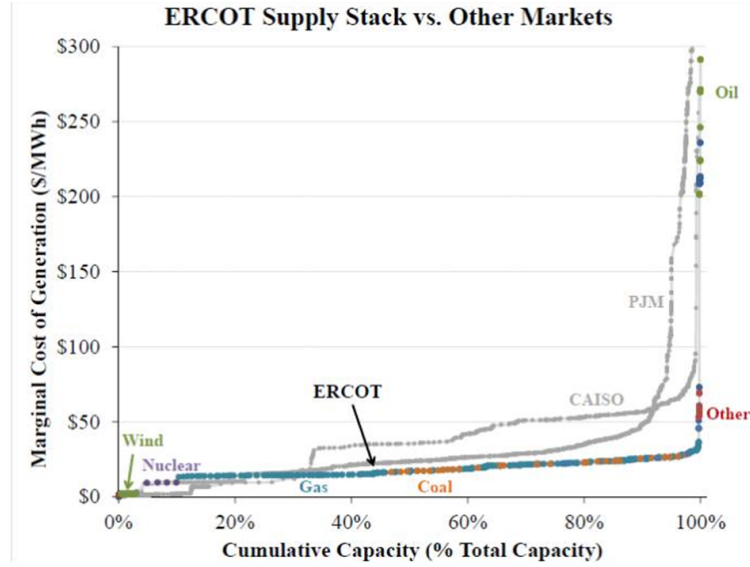


Figure 7.1: Generation-cost curves of three wholesale markets: ERCOT, PJM, CAISO. Source: Brattle Report, pg. 18 <http://www.ercot.com/content/news/presentations/2013/Brattle%20ERCOT%20Resource%20Adequacy%20Review%20-%202012-06-01.pdf>

strates how increased generation results in more sharply increasing costs. Typical daily customer demand has peaks, which thus result in high costs. One of the main goals of DSM is reducing these peaks by *flattening* customer demand throughout the day [96]. Demand-flattening (also called peak-flattening) can reduce both generation costs, infrastructure costs, and  $CO_2$  emissions.

TOU tariffs, which specify different prices for different times of day, were proposed for implementing DSM. Here we define a TOU tariff  $T$  to be a tuple  $T := \langle p_0, p_1, \dots, p_{23} \rangle$ , where  $p_i$  is the electricity price in cents/kWh during hour-of-day  $i$ . We refer to  $p_i$  as *hourly rate*. A TOU tariff with varying hourly rates incentivizes customers to adapt their consumption away from times of peak demand in order to reduce their electricity costs.

### 7.1.2 DSM in Power TAC

In Power TAC [87] as in real-world markets [1], when a tariff is published to the market, customers respond in 3 ways.<sup>1</sup> Responses 1 and 2 take place for any (fixed-rate or non-fixed-rate) tariff publication, while response 3 takes place for non-fixed-rate tariffs, such as TOU, as follows:

1. *subscription changes*: a portion of the customer population may change their tariff subscriptions.
2. *consumption elasticity*: customers elastically adapt their *total* consumption based on prices.
3. *consumption shifting*: customers may shift consumption from expensive to cheap hours.

In competitive retail markets, TOU tariffs may need to compete with *fixed-rate* tariffs, which sell energy for a fixed price per unit. Fixed-rate tariffs do not affect customers' comfort, since customer payments are determined solely by the total energy consumed, regardless of *when* it is consumed. In contrast, under TOU tariffs customers face a trade-off between cost and comfort: to save costs, they may need to change their consumption patterns. Customers will subscribe to a TOU tariff and change consumption if the potential cost saving compared with competing fixed-rate tariffs is large enough to compensate for the extra discomfort.

In Power TAC, about 90% of the consumption is done by factored-customer models, which represent populations of customers (See Section 2.2.2). Factored-customers model the cost-comfort trade-off as follows [87]. A customer has a default *energy profile*  $e_H$ , which is a vector of desired consumption values up to some horizon

---

<sup>1</sup>Since Power TAC customers are autonomous agents representing their human owners, they respond at a higher frequency than a typical human. This setup is expected to reflect future markets with high penetration of home automation systems, and with autonomous agents optimizing residential consumption.

$H$ . Let  $\bar{e}_H$  be a modified energy profile defined by some *admissible permutation* of  $e_H$ . Intuitively, an admissible permutation is a modified energy profile that satisfies the customer's constraints on how energy can be shifted, for instance not consuming below a customer's minimum required demand at any time, and shifting only portions of demand that are flexible. The discomfort implied by an admissible permutation  $\bar{e}_H$  is quantified using a distance metric defined on profile vectors:  $d(e_H, \bar{e}_H)$ . Power TAC currently uses the  $L_2$  distance metric  $d(e_H, \bar{e}_H) := \sum_{t=1:H} (e_t - \bar{e}_t)^2$ , and we find that it has desirable strategic effects, which we elaborate on later. Let  $cost(T, \bar{e}_H)$  be the cost paid by a customer consuming energy according to  $\bar{e}_H$  under a tariff  $T$ . Let  $w$  be a constant weighting the importance of cost vs. discomfort. Then the customer's utility of subscribing to tariff  $T$  and consuming according to  $\bar{e}_H$  is  $u_{cust}(T, \bar{e}_H) := -(cost(T, \bar{e}_H) + w \times d(e_H, \bar{e}_H))$ .

Customers optimize  $\bar{e}_H$  to maximize their utility under a given tariff to which they are subscribed. This formulation assumes a baseline maximum utility of 0 corresponding to the customer using energy for free and consuming according to its desired energy profile. For a customer  $cust$  subscribed to a tariff  $T$ , the optimal consumption profile is  $e_H^* := \arg \max_{\bar{e}_H} u_{cust}(T, \bar{e}_H)$ , and the utility of tariff  $T$  as (overloading notation)  $u_{cust}(T) := u_{cust}(T, e_H^*)$ . For any fixed-rate tariff  $T_{fixed}$ , all permutations have the same price, so by the above definitions  $e_H^* = e_H$ , and  $u_{cust}(T_{fixed}) := -cost(T_{fixed}, e_H)$ . Therefore, for a given TOU tariff  $T_{tou}$  and a fixed-rate tariff  $T_{fixed}$ , the utility of  $T_{tou}$  for a customer  $cust$  is higher than that of  $T_{fixed}$  (i.e.  $u_{cust}(T_{tou}) > u_{cust}(T_{fixed})$ ), when  $-(cost(T_{tou}, e_H^*) + w \times d(e_H, e_H^*)) > -cost(T_{fixed}, e_H)$ , i.e. when it saves enough cost to overcome the extra discomfort.

## 7.2 Our Contribution to DSM in Power TAC

Section 7.1.2 described how Power TAC's factored-customer consumers currently respond to tariff publications and specifically to DSM using TOU tariffs. Part of

this implementation was contributed by us, as follows. Until early in the year 2015, the consumers’ demand-shifting code was disabled in the Power TAC simulator. This code was implemented as described in [87] and could work in a restricted setup, but needed several enhancements to work in a full game in the way described in the previous section. We enhanced the code to work in a full game, as follows.

First, consumers originally evaluated the utility of candidate tariffs using their *default* consumption profile, thus implicitly assuming that they would not shift their consumption. We enhanced consumers to evaluate each TOU tariff using their utility-maximizing shifted consumption profile under this tariff. To find a utility-maximizing shifted consumption profile, customers use the same code that they use for shifting consumption under an actual submission.

Second, customers originally treated all TOU tariffs as equally uncomfortable (using a fixed constant to account for the discomfort of *any* TOU tariff). However, a TOU tariff that requires small consumption-shifting should be considered less uncomfortable than a TOU tariff that requires large consumption-shifting. We enhanced consumers to account for a TOU tariff’s discomfort based on the amount of shifting in their utility-maximizing shifted profile under this tariff. The amount of shifting is measured as the  $L_2$  distance between a consumer’s desired and shifted consumption profiles, as was described in Section 7.1.2.

Third, factored-customer models that represent large populations can subscribe subsets of their populations to different tariffs. The original implementation of demand-shifting found a *single* shifted consumption profile per factored-customer, which maximized utility over *all* tariffs subscribed by the population. As a result, different population members consumed according to the same consumption profile, even if they were subscribed to radically different TOU tariffs. We enhanced factored-customers to allow different population members to have different consumption profiles, where each profile is optimized based on a single tariff to which the



member is subscribed.

Finally, we enabled the corrected demand-shifting for about 50,000 residential customers and 25 office buildings in the simulation, tested it thoroughly, and fixed a few small issues to achieve the behavior described in Section 7.1.2.

### 7.3 Challenges of Using TOU Tariffs in Competitive Markets

Section 7.1.1 stated that TOU could help to flatten demand, and thus reduce infrastructure costs, generation costs, and  $CO_2$  emissions. While reduced-costs and emissions could increase social welfare, it is interesting to ask whether TOU tariffs can benefit a self-interested autonomous broker agent, and explore the broker's incentives to use them. This section characterizes the incentives and challenges in using TOU by a broker in a competitive market.

Since TOU tariffs are less attractive for customers than fixed-rate tariffs, they are not expected to increase the broker's revenue; instead, they can benefit the broker by reducing its costs, thus making the broker more competitive so that the broker could potentially increase its profit. Therefore, there are two basic conditions that need to be met for a TOU tariff to be more beneficial for a broker than the best fixed-rate tariff:

**Condition 1** TOU tariff should reduce the broker's expected costs compared with the best fixed-rate tariff this broker could publish.

**Condition 2** TOU tariff's reduced costs should result in larger expected profit than the best fixed-rate tariff this broker could publish.

Note that the second condition is necessary, since a broker might design TOU tariffs that reduce costs *and* revenue, such that the profit does not necessarily increase compared with the best fixed-rate tariff.

In monopolistic retail markets the surplus resulting from reduced wholesale electricity costs directly benefits the monopoly and possibly the customers. In contrast, in competitive markets this surplus might benefit the competitors (even if they do not use TOU tariffs), since wholesale electricity costs are typically a function of the *total* quantity bought, due to the wholesale auction structure. As a result, brokers using fixed-rate tariffs can enjoy the reduced prices resulting from peak-flattening by another broker using TOU tariffs, while at the same time gaining market share from this TOU broker due to the extra discomfort that TOU tariffs incur on customers.

This chapter focuses on two questions. First, how should an autonomous broker optimize TOU tariffs that are both 1) attractive to customers in a competitive retail market with fixed-rate tariffs and 2) more profitable for the broker than the best fixed-rate tariffs? Second, what is the economic impact of TOU tariffs in a competitive market?

## 7.4 LATTE-TOU

This section views LATTE as executing a black-box optimization, and describes our instantiation of LATTE that uses TOU tariffs, called LATTE-TOU.

### 7.4.1 LATTE as a Black-Box Optimization

LATTE can be viewed as executing a *black-box* optimization problem at every step: the inputs to this optimization problem are tariff and wholesale action combinations; the optimized objective is the predicted utility over the lookahead horizon; and an objective evaluation is computed using a lookahead trajectory that uses the transition and reward function models, implemented in the functions `predictTariffEffects` and `PredictWholesalePrice`. The black-box refers to the lookahead computation that returns a predicted utility for a given action-combination input.

In all of our instantiations, an action-combination input to LATTE’s black-box evaluation is completely determined by the tariff action in the combination: since our LATTE instantiations aim for zero supply-demand imbalance, wholesale actions (quantities to procure) are determined to be the predicted demand for the candidate tariff action. Therefore, our instantiations of LATTE’s black-box optimization is actually performed over candidate tariff actions, which are determined by the function `ComputeNextCandidateTariffAction`.

In our previous instantiations of LATTE (Algorithm 1 in Chapter 4), the function `ComputeNextCandidateTariffAction` generated a set of fixed-rate candidate tariffs (line 5 of LATTE), which were evaluated (along with the corresponding wholesale actions) by LATTE’s black-box optimization (lines 6-24 of LATTE) to determine the best tariff action to execute. Here we would like to extend LATTE’s lookahead to search over TOU tariff actions in addition to fixed-rate tariffs.

From an optimization perspective, TOU tariffs present at least three challenges. First, optimizing a TOU tariff requires searching over a multi-dimensional space: while a fixed-rate tariff has only one continuous price to optimize, a TOU tariff is composed of 24 continuous prices. Moreover, these 24 prices must be optimized in conjunction: changing a single price can affect customer-demand throughout the lookahead horizon due to subscription changes and demand-shifting effects (rather than changing demand just in the timeslot for which this price is effective). Second, a broker can only evaluate a limited number of candidates, due to a real-time constraint on computation. Third, this 24-dimensional optimization is generally intractable: effects such as subscription-changes and consumption-shifting create discontinuities and local maxima in the predicted-utility objective, so that finding a global optimum is generally intractable.

### 7.4.2 LATTE-TOU

Due to the real-time constraint on computation, an instantiation of LATTE that uses TOU tariffs needs to efficiently and effectively sample 24-dimensional TOU tariff candidates to evaluate (addressing challenge (i) from Chapter 3). Due to intractability, we resort to finding a local optimum. Our instantiation of LATTE that uses TOU tariffs is called LATTE-TOU. LATTE-TOU is mostly identical to TacTex-15’s instantiation of LATTE, with two exceptions. First, it can reason about TOU tariffs by using the shifting consumers models (described in Section 7.1.2) to predict customer subscription changes and future consumption. Second, it introduces a new instantiation of the function `ComputeNextCandidateTariffAction`, which generates candidate TOU tariffs based on a local search that uses LATTE’s black-box evaluation.

Before describing the instantiation of `ComputeNextCandidateTariffAction` used by LATTE-TOU, we note that we tried other instantiations by plugging-in well-known local search methods such as Amoeba, BOBYQA, and Powell’s method. All these methods failed to converge under the real-time constraint on computation, and resulted in ineffective TOU tariffs published to the market. Therefore, we designed an empirical gradient-ascent algorithm, which efficiently found a local optimum in our experiments.

Algorithm 6 describes the instantiation of `ComputeNextCandidateTariffAction` as a gradient-ascent algorithm, used by LATTE-TOU. Line 1 maintains a counter counting the number of calls to `ComputeNextCandidateTariffAction`. The algorithm has three phases, where the current phase is determined by the number of calls to the function. Phase 1 (lines 2-4) returns fixed-rate candidates similarly to TacTex-13’s and TacTex-15’s instantiations of LATTE. The goal of this phase is to find the best fixed-rate candidate to serve as a seed for the local TOU optimization phase. Phase 2 (lines 5-20) returns candidates for computing a 2-sided, empirical gradient. Lines

---

**Algorithm 6** LATTE-TOU::ComputeNextCandidateTariffAction()

---

▷ *numCalls* counts the number of calls to this function

1: *numCalls*  $\leftarrow$  *numCalls* + 1

▷ **Phase 1:** return fixed-rate tariffs, like TacTex-13 and TacTex-15

2: **if** *numCalls* < *numFixedRateCandidates* **then**

3:     *fixedRateCandidate*  $\leftarrow$  TacTex-13::ComputeNextCandidateTariffAction()

4:     **return** *fixedRateCandidate*

▷ **Phase 2:** the next 48 calls generate candidates for 2-sided gradient estimation

5: **if** *numCalls* == *numFixedRateCandidates* **then**

6:     *fixedRateSeed*  $\leftarrow$  FindBestFixedRateTariff()

7:      $\langle p, p, \dots, p \rangle \leftarrow \text{ConvertToTOUTariff}(\text{fixedRateSeed})$      ▷ vector of length 24

8:     *numTOUCalls*  $\leftarrow$  *numCalls* − *numFixedRateCandidates*

9:     **if** *numTOUCalls* < 48 **then**

10:         *i*  $\leftarrow \lfloor \frac{\text{numTOUCalls}}{2} \rfloor$

11:         **if** *numTOUCalls* is even **then**

12:             **return**  $\langle p, \dots, p, p + \epsilon, p, \dots, p \rangle$      ▷  $\epsilon$  added to *i*'th entry

13:         **else**

14:             **return**  $\langle p, \dots, p, p - \epsilon, p, \dots, p \rangle$      ▷  $\epsilon$  subtracted from *i*'th entry

▷ Estimate 2-sided empirical gradient (once)

15: **if** *numTOUCalls* == 48 **then**

16:     **for** *i* in 1, ..., 24 **do**

17:          $u_i^+ \leftarrow \text{RetrieveUtility}(\langle p, \dots, p, p + \epsilon, p, \dots, p \rangle)$      ▷  $\epsilon$  added to *i*'th entry

18:          $u_i^- \leftarrow \text{RetrieveUtility}(\langle p, \dots, p, p - \epsilon, p, \dots, p \rangle)$      ▷  $\epsilon$  subtracted from *i*'th entry

19:          $\langle \epsilon_1, \epsilon_2, \dots, \epsilon_{24} \rangle \leftarrow \text{NormalizeGradient}(\langle \frac{u_1^+ - u_1^-}{2}, \frac{u_2^+ - u_2^-}{2}, \dots, \frac{u_{24}^+ - u_{24}^-}{2} \rangle)$

20:          $P \leftarrow \langle p, p, \dots, p \rangle$

▷ **Phase 3:** walk along gradient

21:  $P \leftarrow P + \langle \epsilon_1, \epsilon_2, \dots, \epsilon_{24} \rangle$

22: **return**  $P$

---

5-7 find the fixed-rate candidate with the highest utility (as computed and stored in line 24 of `LATTE`), and use it as a seed for the local optimization, converting it into a TOU tariff with 24 identical prices. Lines 9-14 generate one of the 48 candidates for gradient-estimation, by perturbing each of the 24 prices by  $\pm\epsilon$  (in our experiments  $\epsilon = 0.5\text{cents}/kWh$ , selected based on informal testing to be of a similar scale to price changes during a game). Next, a 2-sided gradient is computed from the 48 perturbations and is normalized to a length of  $\epsilon$  (lines 15-19). In line 20, we initialize the current point in the gradient-ascent  $P$  to the fixed-rate seed. In phase 3 (lines 21-22), each following candidate is generated by taking a step along the computed gradient. This phase is ended externally, by `LATTE`'s `Done()` function, either after a local minimum has been reached, or when time is up.

Ideally, we would compute a new gradient before every step taken during phase 3. However, the real-time constraint on computation time prevents us from doing so: it allows us to evaluate a total of about 70 candidate tariffs, while each gradient estimation requires 48 evaluations. A more sample-efficient gradient estimation method such as the policy gradient employed by [51] could be explored in the future. Due to the limitation of about 70 evaluations we also had to reduce the number of evaluated fixed-rate candidates: we did it by using a binary-search over fixed-rate prices, so that instead of evaluating all 100 candidates generated by `TacTex-13`'s instantiation, we evaluated about 8 fixed-rate candidates.

## 7.5 Results

We evaluated our TOU broker which uses `LATTE-TOU` using paired tests. We measured the impact of modifying a component of the broker by testing the original and the modified version in a set of games, in which the opponents and most random factors in the simulation were held fixed (random seeds, weather conditions). Paired testing improves our ability to evaluate the statistical significance of the results,

by allowing us to use the Wilcoxon matched-pairs signed-ranks test instead of an unpaired test. To fix weather conditions, we used weather files containing 3 months of real-world weather. To cover year-round weather conditions we used 8 weather files (each file used by 1/8 of the games) with start-dates of January, April, July, October of 2009 and 2010. Opponents were fixed to be one of the top brokers played in the Power TAC 2014 finals: AgentUDE (1st place), and CWIBroker (2nd place). The simulator version used in the experiments in this section is specified in Appendix A.

### 7.5.1 Impact of LATTE-TOU on Broker’s Performance and on the Economy

We tested how using TOU tariffs optimized with LATTE-TOU affected 1) the broker’s performance. and 2) the economy. We compared our TOU Broker which uses LATTE-TOU with two variations: one that uses fixed-rate tariffs and another that uses a naive TOU tariff optimization. We refer to these brokers as TacTex-TOU, FixedRate, and TOUNaive. FixedRate was created from TacTex-TOU by disabling phase 2 and 3 of `ComputeNextCandidateTariffAction`, and using the fixed-rate tariff returned by phase 1 (line 1). TOUNaive was created from TacTex-TOU as follows. Phase 2 and 3 of `ComputeNextCandidateTariffAction` were replaced with a phase that naively assigns higher rates to hours with higher predicted costs, by adding a fixed margin to these predicted costs. Specifically, given a fixed-rate tariff with rate  $p$  returned by phase 1, and given a predicted cost vector for the next 24 hours  $(c_{+1}, \dots, c_{+24})$ , the naive algorithm computes an average margin  $m := \frac{1}{24} \sum_{i=+1}^{+24} (p - c_i)$ , and publishes a TOU tariff with the price vector  $P = (c_{+1} + m, \dots, c_{+24} + m)$ . All other broker components remained identical between the three brokers. We compared these three brokers in 2 different experiments of 200 games each: (1) playing against AgentUDE and (2) playing against

CWIBroker, both of which use only fixed-rate tariffs.

Table 7.1 shows the results of these two experiments. Each row shows a measured quantity averaged over games played by `FixedRate`, `TOUNaive` and `TacTex-TOU`, as well as the relative change in this quantity when using `TacTex-TOU` instead of `FixedRate`. All results are statistically significant with  $p = 0.01$  (many with  $p \ll 0.01$ ), using the Wilcoxon matched-pairs signed-ranks test. `TacTex-TOU` was the only agent that earned a higher score than both competitors (by 9% and 70% gaps against `CWIBroker` and `AgentUDE` respectively). `TOUNaive` was dominated by `TacTex-TOU` in the sense that it made less profit against both opponents, and lost to `CWIBroker`.

Compared with `FixedRate`, `TacTex-TOU` either earned more profit (against `AgentUDE`), or increased its profit-share from losing to winning (against `CWIBroker`, although with lower profits), while reducing peak demand by around 15%. `TacTex-TOU`'s peak reduction reduced the electricity costs for both brokers and customers (including competitors'), and therefore increased social welfare. The surplus resulting from peak-reduction benefited either brokers or customers, depending on broker strategies. When playing against `CWIBroker`, customers enjoyed a 9.5% cost reduction, and brokers' suffered profit reduction, due to a fierce price-reduction competition. On the other hand, when playing against `AgentUDE`, brokers did not reduce prices as much; customers' cost reduction was only 1%, while brokers' profits increased. Since our broker plays a best-response strategy, the difference depends on how cooperative the other broker is.

Table 7.2 shows the results of running `TacTex-TOU` against itself. It achieved the best flattening (around 20% and 5% peak reduction compared with `FixedRate` and `TacTex-TOU`), and the lowest price for customers (around 5%-20% savings compared with `FixedRate` and `TacTex-TOU`). However, in this case `TacTex-TOU` achieved the lowest profit of all brokers due to a fierce price-reduction competition.



Table 7.1: FixedRate, TOUNaive and TacTex-TOU competing in 2-broker games against AgentUDE, CWIBroker (TacTex-TOU is denoted TOU).

	(a) Fixed-Rate-vs-UDE	(b) TOUNaive-vs-UDE	(c) TOU-vs-UDE	Change (c)/(a)
score: our-agent (M\$)	<b>1.893</b>	<b>1.689</b>	<b>1.922</b>	1.016 (+1.6%)
score: UDE (M\$)	<b>0.895</b>	<b>0.578</b>	<b>1.122</b>	1.253 (+25.3%)
market-share: our-agent (%)	64.0	73.3	61.4	0.959 (-4.1%)
(our) avg electricity-buy price	0.053	0.051	0.051	0.963 (-3.7%)
(our) avg electricity-sell price	0.105	0.098	0.105	1.000 (-0.0%)
(all) avg electricity-buy price	0.051	0.049	0.049	0.961 (-3.9%)
(all) avg electricity-sell price	0.105	0.099	0.104	0.990 (-1.0%)
peak-demand (MW)	<b>86.771</b>	<b>71.882</b>	<b>73.519</b>	0.847 (-15.3%)

	(a) Fixed-Rate-vs-CWI	(b) TOUNaive-vs-CWI	(c) TOU-vs-CWI	Change: (c)/(a)
score: our-agent (M\$)	<b>0.677</b>	<b>0.524</b>	<b>0.622</b>	0.919 (-8.1%)
score: CWI (M\$)	<b>0.771</b>	<b>0.620</b>	<b>0.558</b>	0.724 (-27.6%)
market-share: our-agent (%)	44.2	54.3	54.7	1.238 (+23.8%)
(our) avg electricity-buy price	0.057	0.054	0.054	0.947 (-5.3%)
(our) avg electricity-sell price	0.095	0.087	0.086	0.905 (-9.5%)
(all) avg electricity-buy price	0.057	0.055	0.053	0.930 (-7.0%)
(all) avg electricity-sell price	0.094	0.086	0.086	0.915 (-8.5%)
peak-demand (MW)	<b>86.701</b>	<b>74.720</b>	<b>73.651</b>	0.849 (-15.1%)

Table 7.2: Self-play (TacTex-TOU vs TacTex-TOU), compared with TacTex-TOU vs AgentUDE, and TacTex-TOU vs CWIBroker (TacTex-TOU is denoted TOU).

	(d) TOU-vs-TOU	Change (d)/(c) (UDE)	Change (d)/(c) (CWI)
score: our-agent (M\$)	<b>0.493</b>	0.257 (-74.3%)	0.791 (-20.9%)
score: agent-copy (M\$)	<b>0.482</b>	—	—
market-share: our-agent (%)	50.5	0.823 (-17.7%)	0.927 (-7.3%)
(our) avg electricity-buy price	0.051	1.000 (-0.0%)	0.944 (-5.6%)
(our) avg electricity-sell price	0.083	0.790 (-21.0%)	0.954 (-4.6%)
(all) avg electricity-buy price	0.051	1.041 (+4.1%)	0.944 (-5.6%)
(all) avg electricity-sell price	0.083	0.798 (-20.2%)	0.954 (-4.6%)
peak-demand (MW)	<b>70.101</b>	0.954 (-4.6%)	0.947 (-5.3%)

TacTex-TOU’s best-response self-play benefited customers but not the broker. This illustrates game-theoretic considerations pointed out by [63], whereby cooperative brokers could make higher profits, in this case by enjoying more of the surplus created by peak-reduction, at the expense of customers.

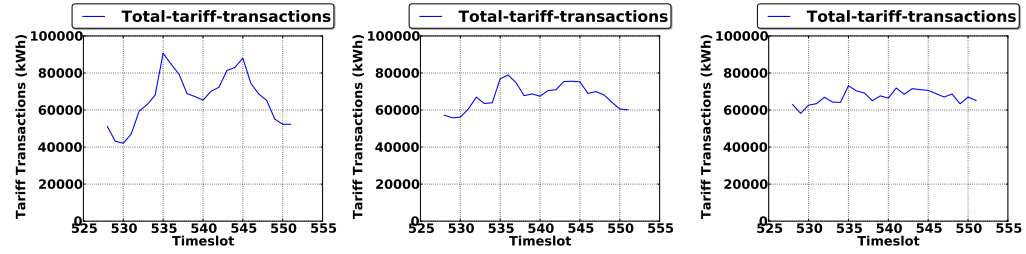
Figure 7.2 shows how the market power of a TOU broker affects its ability to flatten demand. All plots show consumption over 24 simulated hours. The left, middle, right columns show FixedRate, TOUNaive and TacTex-TOU playing against CWIBroker. Peak demand is around 90MW, 80MW, and 70MW respectively (top row). TacTex-TOU’s large market share allowed it to counter-balance CWIBroker’s customers’ peaked-demand, while TOUNaive was only partially successful in doing so

due to both lower market share (middle row), and suboptimal TOU pricing (bottom row).

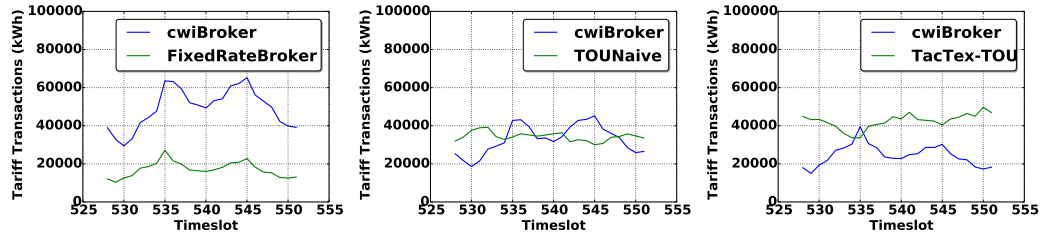
TacTex-TOU’s frequent replanning using LATTE-TOU prevented customer herding (many customers shifting consumption to times with the lowest price, causing a new peak [83]). Even though we disabled two main customer-components for addressing herding (bundle-based optimization and stochastic shifting [87]) and let customers shift greedily to their utility-maximizing electricity profile, no herding was observed, due to a combination of (1) a TOU broker that implicitly coordinated flattening through profit-maximizing tariffs (Figure 7.2c, right), and (2) a smooth discomfort metric  $\text{dist}(e_H, e_H^*)$ . This coordinated flattening underlines a potential benefit of employing TOU brokers in competitive markets.

### 7.5.2 Robustness of TOU to Prediction Errors

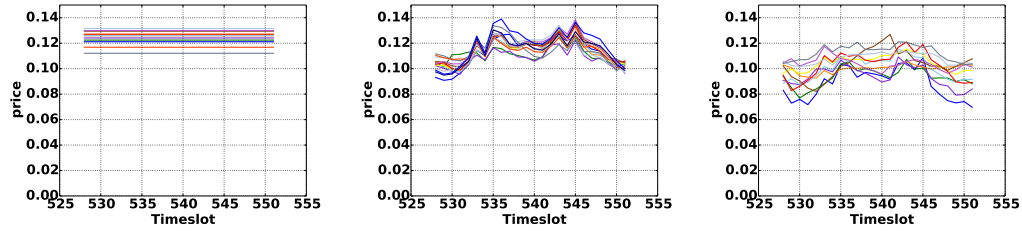
We tested the robustness of TacTex-TOU to errors in its consumption-shifting predictions. Table 7.3 compares profits and peak-demand when testing 2 variations of TacTex-TOU against CWIBroker. We chose CWIBroker as an opponent against which TacTex-TOU had smaller profit margins (see Table 7.1), so accurate predictions seem important. The left column shows the results of TacTex-TOU, copied from Table 7.1, as a reference. The NoShift broker was created from TacTex-TOU by disabling the consumption-shifting prediction module, and the FlatCost broker was created by adding noise to cost-prediction, making it predict a flatter cost-curve slope. Based on Table 7.3, the consumption-shifting prediction module is critical to both earning profits and peak-flattening: without it TacTex-TOU lost to CWIBroker by about a 10% gap, and peak-demand was barely reduced compared to FixedRate. Similarly, when the predicted cost-curve was too flat, TacTex-TOU lost with negative profit and almost no flattening.



(a) Total consumption over 24 hours.



(b) Per-broker consumption over 24 hours.



(c) Active tariffs over 24 hours.

Figure 7.2: Consumption flattening: FixedRate (left column), TOUNaive (middle), TacTex-TOU (right).

Table 7.3: Ablation analysis: erratic-predictions

	<b>TacTex-TOU</b>	<b>NoShift</b>	<b>FlatCost</b>
score: our-agent (M\$)	0.622	0.507	-0.007
score: CWI (M\$)	0.558	0.550	0.210
peak-demand (MW)	73.651	83.728	82.779

## 7.6 Chapter Summary

We formalized the problem of TOU tariff optimization in competitive retail markets, and proposed a real-time gradient-based, utility-optimization (profit-maximization) algorithm that approximates its solution. Our algorithm is fully implemented and tested extensively in the Power TAC simulator. Our gradient algorithm is currently the only TOU algorithm that performs robustly in Power TAC’s complex, realistic environment: both a naive approach (TOUNaive) and well-known optimization algorithms failed to improve upon our fixed-rate broker and/or to outperform the top 2014 brokers. We have shown that TOU tariffs can compete successfully with fixed-rate tariffs: our TOU broker agent outperformed the top 2 broker agents of the Power TAC 2014 finals, reduced peak-demand by 15% compared with using only fixed-rate tariffs, increased its profits and/or profit-share, and saved costs for all customers (including competitors’). Our ablation analysis showed the importance of having accurate customers shifting-predictions and cost-curve predictions.

While TOU tariffs can induce *customer-herding*, our TOU broker prevented it by implicitly coordinating flattening through profit-maximizing tariffs. This coordinated flattening underlines a potential benefit of employing autonomous TOU brokers in competitive power markets. In addition, we have seen that a TOU broker’s customer share is an important factor in its ability to flatten demand: to counter-balance peaked consumption of fixed-rate brokers’ customers, it needs to gain large customer-share by creating attractive TOU tariffs that are still profitable. Finally, our experiments demonstrated game-theoretic issues that affect the distribution of surplus created by reduced costs. An important direction for future work is exploring the market efficiency when many broker agents are competing against each other.

# Chapter 8

## Related Work

This main motivation for this dissertation is exploring how modern artificial intelligence (AI) techniques can contribute to society’s shift towards wide scale deployment of smart grids, which make electricity power distribution more efficient and flexible. As such, this dissertation makes contributions within both the general field of AI, and to the literature on smart grids and electricity markets. Within AI, it is particularly situated with in the subfield of multiagent systems (MAS). Figure 8.1 illustrates how Power TAC, the substrate domain of this dissertation, fits into these research areas: it lies at their intersection, and inside the Trading Agent Competitions (TAC) area, which lies inside the Agent-based Computational Economics (ACE) area, which lies inside MAS. This section reviews research that is most related to this dissertation within these fields.

Before reviewing related work, we note that this dissertation also contributes to the broad field of *computational sustainability*. The call for starting a new research area of computational sustainability was made by Gomes [23]. In this paper, the author described the need to harness computational resources to address environmental issues, including management of natural resources such as energy, for the benefit of current and future generations. Power TAC’s research goals include using

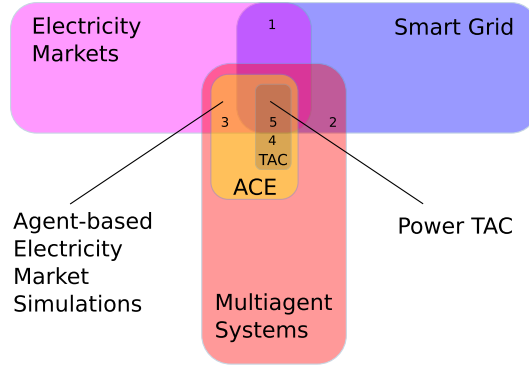


Figure 8.1: **Dissertation contribution areas.**

autonomous agents for the benefit of sustainable energy, and therefore, in addition to the areas shown in Figure 1, it also falls within the area of computational sustainability. The literature on computational energy sustainability has been growing rapidly and includes applications such as smart home heating, micro-storage management, wind forecasting, wind turbine optimization, energy disaggregation, solar tracking, electric vehicle charging and others [89, 122, 121, 40, 119, 21, 128, 61, 60, 59, 38, 129, 54, 53, 55, 52, 70, 41, 117].

In the rest of this section we overview related work based on the numbered regions in Figure 8.1. Specifically, we overview research in smart grid and electricity markets, with a special focus on Time-Of-Use tariffs and demand-side management (Region 1, Section 8.1); MAS applications for the smart grid (Region 2, Section 8.2); Agent-based Computational Economics (ACE) in the context of electricity markets (Region 3, Section 8.3); other Trading Agent Competitions (TAC) (Region 4, Section 8.4); and Power TAC (Region 5, Section 8.5).

## 8.1 Smart Grid and Electricity Markets

The need for a smart grid and its potential benefits have been laid out in official policy documents [114, 108] as well as academic research (e.g. [2]). Basic background

on power market economics can be found in the book *Power System Economics* [99]. Electricity markets are going through a major transition from traditional, regulated monopolies into deregulated, competitive markets [35]. At the same time, customers are being engaged in power markets, to incentivize flexible demand that adapts to supply conditions [114]. These major changes create new challenges in power markets design. In the context of these challenges, we have especially focused on Time-Of-Use tariffs (Chapter 7). We survey related work in this area next.

### 8.1.1 Time-Of-Use Tariffs and Demand-Side Management

Adapting electricity demand to supply conditions can be beneficial from different reasons, some of them were listed back in the 1980s [96]. An analysis of the California Energy Crisis concluded that the risk of such a crisis could be greatly mitigated by customer demand that responds to electricity prices [7]. Demand-side management (DSM) is viewed as an important component of future smart grids [114]. Until recently, the technological infrastructure that is needed for implementing DSM was missing, but with recent advances, such as smart-meter installations at customer homes, different forms of DSM are expected to be implemented in the relatively near future. A taxonomy of DSM is provided by Palensky et al. [68], which additionally overviews several demonstration projects. DSM can be implemented through (a) *direct load control* methods (e.g. [8]), in which customers give electric utilities direct control over their devices and the option to turn them off to reduce load; or (b) *indirect methods*, such as pricing-based incentives.

In recent years, two types of pricing-based DSM have attracted growing attention both in academia and in industry: Time-Of-Use (TOU) tariffs, and Real-Time-Pricing (RTP) tariffs [130, 15, 16, 127, 10, 95, 65, 11]. TOU tariffs were used for a long time in various countries for large customers, e.g. by Electricite de France as far back as 1956 [12]. With recent advances, such as demand-side smart-

metering for residential customers, TOU pricing schemes have been considered for residential use. For example, as of the current date, residential TOU pricing schemes are proposed to take effect in California starting January 2019 [91]. Existing work on TOU tariffs either has not considered competitive retail markets or has used more abstract, smaller-scale simulations compared with Power TAC [130, 127, 10, 14, 118, 1, 107, 131, 19]. In Power TAC, to the best of our knowledge the first broker that used TOU tariffs was Mertacor13 [66] (see Section 8.5). Mertacor13 used TOU tariffs with 2 or 3 daily rates. However, at that time the Power TAC simulator included only non-shifting customers, so that the impact of TOU tariffs in presence of demand-shifting customers could not be tested. Other than Mertacor13, Power TAC broker agents used fixed-rate tariffs [67, 31, 111, 63, 4, 56, 57].

Therefore, to the best of our knowledge, our research is the first to investigate the usage of TOU tariffs by autonomous brokers in a large-scale, detailed, realistic simulation of competitive retail power markets with autonomous, demand-shifting customer agents.

The issue of free-riding customers was described in several references. Free-riding customers are customers who benefit from reduced costs due to demand-shifting of other customers without changing their own demand. According to Holland et al. [30], such customers could attain up to 90% of the benefit of RTP (real-time pricing) adoption by other customers. Other research analyzed the potential savings of shifting customers and free-riders as a function of the percentage of shifted quantity [98]. Some solutions that discourage free-riding were offered by Horowitz et al. [32]. These solutions aim at rolling the savings due to demand-shifting more fairly on customers, based on each customer’s contribution to peak-flattening. However, their proposed methods use hypothetical, non-shifted customer consumption profiles as a basis for their computation. They offer to construct such profiles from historical data, but this may be challenging to do over the long run, since hypo-



thetical, non-shifted consumption patterns may change over time without being observed (since only the actual, potentially shifted, consumption can be observed). In Power TAC's competitive retail market, free-riding can occur among fixed-rate brokers who benefit from peak-reduction of non-fixed-rate brokers, and thus can offer their customers attractive fixed-rate tariffs.

## 8.2 MAS for the Smart Grid

The need and the potential benefits of employing artificial intelligence (AI) for the benefit of the smart grid have been laid out by Ramchurn et al. [84]. In this paper, the authors have argued that the smart grid vision, as laid out by the U.S. Department of Energy [114], not only presents challenges for power systems engineering, telecommunications and cyber-security, but also embodies concepts that have long been investigated in the computer science and AI communities, such as distributed intelligence, automation and information exchange. The authors have argued that the smart grid provides new challenges to be solved by the AI community. Therefore, they have proposed the foundations of a new research agenda of using AI for solving challenges of the smart grid. Some of these challenges require algorithms that solve problems involving large number of heterogeneous, self-interested participants (such as different types of consumers and generation facilities), under high-levels of uncertainty and dynamism. The authors have highlighted how such issues appear in key components of the future smart grid: demand-side management, electric vehicles, virtual power plants, the emergence of prosumers (customers that both consume and produce energy), and self-healing networks. Power TAC addresses many of these challenges by providing a rich, multiagent simulation environment that includes prosumers (consumers and renewable producers), demand-side management capabilities (both pricing based and direct load-control based), and electric vehicles. Power TAC investigates economically motivated decisions of large number of actors,

which are at the core of the challenges laid out by Ramchurn et al. [45].

In the recent years multiagent systems were proposed to address different challenges in the smart grid. Some examples follow next. A multiagent system was proposed for managing micro-grid operations [78]; this system facilitated seamless transition of the micro-grid from a grid-connected state into an isolated island mode, upon detection of upstream outages. This type of research is complementary to Power TAC, since it models the distribution network in more detail than Power TAC, but lacks Power TAC's large-scale, detailed customer models.

Vytelingum et al. explored the theoretical and practical foundations of agent-based micro storage implementation in the smart grid [123]. They developed a game-theoretic framework to analyze strategic choices made by agents controlling micro-storage in the smart grid, and devised a novel micro-storage strategy that allowed an agent to maximize its owner's savings. This dissertation focuses on a different problem of designing a broker agent that acts effectively in a future smart grid electricity markets in the presence of autonomous customer agents. However, this research can complement Power TAC, since such storage-controlling agents could be incorporated into the Power TAC simulator's customer models.

Ramchurn et al. proposed an agent-based system for decentralized demand-side management [83]. Their research focused on using agents for managing customers' demand. Therefore this research can be incorporated into Power TAC's customer models. Their experiments showed a *herding* phenomenon, which occurred when many customer agents shifted their consumption to low-price times, thus creating new peaks. This dissertation focuses on a related but different topic of designing broker agents that operate in presence of autonomous customers, and shows that autonomous utility-maximizing brokers can prevent herding, by implicitly coordinating customers' consumption through Time-Of-Use (TOU) tariffs that are designed to optimize the brokers' profits.

Agent-based systems were proposed for coordinating energy procurement by a customer-cooperative [118]. This research focused on the customers' perspective, rather than on the retail brokers' perspective which is the focus of this dissertation: they assumed that a coordinator represented a consumer cooperative in procurement of energy for minimal costs. This coordinator has not set energy prices or aimed to incur profits by selling electricity, as our autonomous broker agent do.

Agent-based systems were proposed for automating energy savings in buildings [90, 109, 110, 58, 69]. Such research could be incorporated into Power TAC's customer models, by modeling different aspects of customer energy-savings behaviors in detail.

Recent publications related to Power TAC described novel customer agent models such as electric vehicles that optimized charging based on prices [116], electric vehicle based Virtual Power Plants (VPP) [36, 37], and negotiated learning for autonomous customer agents [88]. Such models are being gradually incorporated into Power TAC.

Other research investigated the use mechanisms and trading strategies for the smart grid and modeled smart grid operations [120, 119]. Related research investigated autonomous trading brokers [85, 86, 77] (described in more detail in section 8.5). While the results of this research were promising, as was pointed out, this research was limited in two important ways [45]: (a) limited scope, and/or (b) limited competitiveness and comparability. In contrast, a major benefit of Power TAC is that it is an open platform that is available as a test bed and a benchmark for any research group.

### 8.3 Agent-Based Power Market Simulations

The complexity of electricity markets calls for rich modeling techniques that can help to understand and analyze their dynamics. Traditional modeling methods are usu-

ally not able to capture the complex dynamics of electricity markets. Agent-Based Computational Economics (ACE) [106] is a computational study of economies modeled as evolving systems of autonomous interacting agents. ACE can offer methods for realistic electricity market modeling. Indeed, different electricity market simulations have been developed using the ACE paradigm (e.g. AMES [62], NEMSIM [28], EMCAS [13], MASCEM [81], and others), many of which focus on wholesale market modeling. Surveys of these simulations and many others were done by Weidlich et al. and Zhou et al. [125, 132].

Power TAC extends the ACE paradigm by creating a rich retail power market simulation and inviting research teams to develop agents that act as retail brokers in the simulation and enter them to an annual competition. Power TAC is therefore a flexible, competitive platform that is easy to use for research and benchmarking purposes by teams from different research groups.

## 8.4 Autonomous Trading Agents and TAC

Autonomous trading agents has been an active research area in the AI community, and there is an immense body of literature in this area. Therefore, we will focus on approaches that are directly related to LATTE, and refer the reader to surveys of approaches that are less directly related.

To the best of our knowledge, the earliest research on autonomous trading agents introduced *shopbots*, which collect information about price and quality of goods, and *pricebots*, which automate the price-setting process for sellers [27, 25, 43, 26, 42].

Power TAC is a specific instance of a Trading Agent Competition (TAC)<sup>1</sup>. Trading agent competitions were held annually since 2000. Trading agent competitions other than Power TAC were in domains such as autonomous bidding for multi-

---

<sup>1</sup><http://tac.sics.se>

ple interacting goods [126], autonomous supply-chain management (TAC-SCM) [94], ad-auctions (TAC-AA) [33], and ad exchange (TAC-Adx) [104].

LATTE is a lookahead policy that optimizes the predicted utility of action combinations. Utility-optimization approaches were used in prior trading agent competitions, however the game setups and the problems they solved, and consequently the methods used, were different than LATTE. For example, decision-theoretic bidding using Monte-Carlo estimation of the clearing price distribution was used for one-sided auctions [102, 101]. A *predictive-planning* approach in the supply-chain management was introduced by Pardoe et al. [72, 73], in which planning and scheduling were executed based on estimations of future resource availability and constraints. Such an approach can be viewed as a lookahead policy over the underlying domain’s MDP (such as LATTE), although it was not described in that way. This approach executed interdependent optimization, optimizing supply inventories and procurement given demand predictions, and optimizing production and sales, given predictions of supply inventories and future deliveries. The TAC-SCM domain has some similarities to Power TAC: in both competitions autonomous trading agents compete for maximizing profit by buying goods from suppliers and selling them to consumers. However, the trading mechanisms used in TAC-SCM were request for quotes, in market structures different then the electricity market structures in Power TAC. Consequently, the specific problems solved and the methods used were different than LATTE.

In TAC-AA, another utility optimization approach was used [71]: the agent operated by estimating the full game state from limited information using methods such as particle filter [75], used these estimates to make predictions, and optimized its actions (daily bids, ads, and spending limits) with respect to these predictions. Descriptions of these methods can be found in Pardoe’s Ph.D. dissertation [76], or in a shorter version [74].

Some other competitors in trading agent competitions used methods other than utility-optimization. A survey of 22 agent strategies used in the first trading agent competition (in 2000) is provided by Stone et al. [100], and a survey of strategies from the 2002 trading agent competition is provided by Greenwald et al. [24]. Examples of other proposed approaches to agent design included a game theoretic analysis of the economy [48] and fuzzy reasoning [29].

## 8.5 Power TAC Broker Agents

The research that is most directly related to ours is that of other Power TAC broker developers. Our work differs from this research along two dimensions: *problem-formulation*, and *broker-strategy*. This section summarizes the differences along each of these dimensions as applicable, first in general and then in more detail.

Along the problem formulation dimension, this dissertation is the first to formalize the *complete* MDP defined by the underlying broker electricity trading problem. Previous research either did not formulate the trading problem explicitly, or used an MDP to model either a more *abstract* trading problem [85, 86], or a *subproblem* of the complete trading problem [77, 57, 111, 56, 4, 67]. Moreover, all these MDP models were *heuristically* and *manually* constructed. In contrast, our MDP is defined by the underlying problem. Along the broker-strategy dimension, related research either did not explicitly optimize the actions’ predicted utility, or used tariff optimization strategies that can be viewed as special cases of LATTE and which were developed in parallel to LATTE. A detailed comparison follows next.

### 8.5.1 Early Power TAC Brokers

To the best of our knowledge, the first work that developed strategies for autonomous electricity-trading retail broker agents was that of [86] and its subsequent [85]. This work modeled a more abstract broker trading problem, which did not include whole-

sale trading, and assumed fixed customer consumption instead of the variable daily load profile of Power TAC customers. Their broker approximated this abstract trading problem as an MDP using manually-constructed state features on top of a state-space discretization, and using a fixed set of 6 manually defined pricing tactics that were used as MDP actions. This broker used a Q-learning [124] based strategy on top of this approximate MDP model. Therefore, along the problem formulation dimension, the work in [86, 85] modeled a more *abstract* trading problem than ours, and used a *manually constructed, heuristic* MDP model instead of the complete MDP of the underlying domain that we use. Along the broker strategy dimension, similarly to our work, this work optimized the predicted expected utility; however it did so over approximate, manually constructed models rather than over the complete, underlying MDP. Their smaller, approximate MDP model allowed for using Q-learning, while in our complete MDP model Q-learning would be computationally intractable.

SELF [77] was a broker strategy developed in an early, simplified simulation environment (compared with Power TAC), in which (a) a small number of 50 customers was used, (b) daily wholesale markets were randomly sampled from past price-data, so that a broker did not impact wholesale prices, (c) any wholesale order of broker was fulfilled, (d) imbalance fee was a fixed-price per-unit (rather than a function of all brokers' imbalances). SELF modeled the tariff selection problem as an MDP using manually constructed state features and a set of 6 pre-defined tariff actions, and used the SARSA RL algorithm [92] with function approximation and feature-selection/regularization to select tariff market actions. Therefore, along the problem formulation dimension, SELF used an MDP to model the tariff strategy *subproblem* of a more *abstract* trading problem, using *manually constructed, heuristic* state and action spaces. Along the broker strategy dimension, similarly to our work, SELF optimized the predicted expected utility; however it did so over ap-

proximate, manually constructed models rather than over the complete, underlying MDP. Their smaller, approximate MDP model allowed for using SARSA, while in our complete MDP model SARSA would be computationally intractable.

### 8.5.2 Power TAC 2013 agents

CwiBroker13 [63] (2nd place, 2013) used two different tariff strategies. For duopoly markets, it used a tariff strategy inspired by Tit-for-Tat. For oligopoly markets it used a tariff strategy that generated candidate fixed-rate tariffs and estimated their future profits. The idea of estimating tariff profits has similarities to LATTE’s utility-optimization, although it seems to have been implemented differently. They report that this oligopoly strategy did not work well, and that a fallback heuristic strategy was responsible for improving its performance in the oligopoly setup. Later versions of CwiBroker (see below) abandoned this profit-estimation based oligopoly strategy and moved to a new heuristic-based strategy. CwiBroker13’s wholesale strategy introduced the idea of multiple bids among Power TAC brokers, but was based on equilibria in continuous auctions, rather than TacTex-15’s hedging between optimistic strategic bidding and truthful bidding. Therefore, along the problem formulation dimension, we are not aware of any MDP models used by CwiBroker13. Along the broker strategy dimension, CwiBroker13’s Tit-for-Tat strategy is different than our utility-estimation based strategy; CwiBroker13’s oligopoly strategy has similarities with our utility-based strategy, but was implemented differently; CwiBroker13’s equilibrium-based wholesale bidding strategy was different than our MDP-based wholesale bidding.

CrocodileAgent13 [4] (4th place, 2013) used a variant Roth-Erev reinforcement learning algorithm to coordinate wholesale bidding across different markets by choosing among one of four pre-implemented wholesale strategies. Therefore, along the problem formulation dimension, CrocodileAgent13 used an MDP to model a



*subproblem* of the complete trading problem, using *manually constructed, heuristic* state and action spaces.

The 2013 AstonTAC agent [57, 56] (5th place, 2013) used an MDP to model the wholesale bidding process and a separate SMDP [103] to model the tariff selection process. In its bidding MDP, AstonTAC assumed an underlying discrete model for wholesale clearing prices (HMM), where 20 possible states were built offline from a game’s bootstrap data. It used a discrete set of manually constructed reward values to represent wholesale bidding performance. In its tariff selection SMDP, AstonTAC used manually selected sets of finite states and actions. Therefore, along the problem formulation dimension, AstonTAC used an MDP and an SMDP to model *subproblems* of the complete trading problem, using discrete, *manually constructed, heuristic* state and action spaces. In addition, AstonTAC’s MDP differs from our bidding MDP (Chapter 5), in that (i) our bidding MDP does not assume an underlying model of the market, but rather uses a more flexible, non-parametric model of clearing prices at every state, (ii) our reward is determined by actual prices rather than by a manually constructed set of reward signals. Along the broker-strategy dimension, similarly to our work, AstonTAC optimized the predicted expected utility; however it did so over approximate, manually constructed models rather than over the complete, underlying MDP.

Mertacor13 [66] used two types of tariff strategies: (i) a tariff formulation strategy, and (ii) a tariff update strategy. Both strategies were treated as optimization problems, where the broker’s objectives were both maximizing its profit and maintaining an acceptable customer market share. Mertacor13’s general approach was to create a set of 4-6 dimensional particles, each representing a tariff, and use the predicted broker profit as an objective and a Particle Swarm Optimization (PSO) algorithm to search for a tariff with a highest predicted profit. Mertacor13’s market share affected the particle search space boundaries. The 4-6 particle parameters

represented (i) a signup fee, (ii) a periodic fee, (iii) an early withdraw penalty, and (iv) either one fixed-rate price, or three Time-Of-Use rates specifying different prices for three periods that together cover a 24-hour period. Along the problem formulation dimension, Mertacor13 did not explicitly define an MDP model. Along the broker-strategy dimension, similarly to our approach, Mertacor13’s strategy aimed at optimizing the predicted tariff profits. In contrast to our approach, Mertacor13 (1) generated candidates using the Particle Swarm Optimization algorithm, and (2) did not seem to incorporate into its utility predictions the effect of publishing a tariff on the profits of its existing tariffs.

To the best of our knowledge, Mertacor13 was that first broker that used Time-Of-Use tariffs. In contrast to LATTE-TOU which uses 24 hourly rates and a gradient-ascent optimization algorithm, Mertacor13 used three rates and a PSO optimization algorithm. Mertacor13’s Time-Of-Use tariffs’ performance was reported to be slightly lower than its fixed-rate tariffs’ performance. We hypothesize that this lower performance was due to the fact that customers did not have shifting capabilities at that time, and therefore viewed a Time-Of-Use tariff as having higher discomfort (based on a fixed discomfort factor that was implemented in this earlier version of the Power TAC simulator) without reducing their costs.

### 8.5.3 Power TAC 2014 agents

AgentUDE14 [67] (1st place, 2014) used an empirically tuned, heuristic tariff strategy that bound customers with early withdraw penalties and provoked competitors to reduce prices, so that customers would withdraw and pay withdraw-penalties. In the wholesale market AgentUDE14 used Q-learning. Therefore, along the problem formulation dimension, AgentUDE14 used an MDP to model a *subproblem* of the complete trading problem. Along the broker-strategy dimension, AgentUDE14’s empirically-tuned heuristic strategy differs from our utility-optimization approach.

CwiBroker14 (2nd place, 2014) [31] used tuned heuristics based on domain knowledge: in the wholesale market, it adapted its bids towards bids that would have been cleared in recent auctions, and in the tariff market it reduced prices in a pace that is inversely proportional to its market share. Along the problem formulation dimension, to the best of our knowledge CwiBroker14 did not use an MDP model. Along the broker strategy dimension, CwiBroker14’s strategy differs from our utility-optimization approach in that it did not explicitly optimized a utility measure. A detailed analysis of the 2014 Power TAC finals can be found at Babic et al. [5].

#### **8.5.4 Power TAC 2015**

In our experiments, we have used broker agent binaries that were released after the 2015 Power TAC finals. However, at the time of this writing, we are not aware of any publications describing these 2015 agents: the most recent publications that we know of describe the 2014 agents.

## **8.6 Chapter Summary**

This section provided an overview of related work in the areas of smart grid and electricity markets, multiagent systems for the smart grid, agent-based electricity market simulations, autonomous trading agents inside and outside the Trading Agent Competitions (TAC), and Power TAC. Based on this literature review, this dissertation makes several contributions to the state-of-the-art.

Smart grid electricity markets were studied inside and outside the context of multiagent systems. However, past research has not investigated the dynamics of future retail electricity markets in a large-scale, realistic, detailed simulation such as Power TAC, specifically in presence of real-time smart-metering and autonomous agents acting on behalf of customers and retailers.

Using Power TAC as a substrate domain, this dissertation is the first to for-

malize the complete autonomous electricity trading problem faced by a broker agent in future electricity markets. This dissertation then introduces LATTE. LATTE is a lookahead-policy that optimizes the broker’s predicted utility and approximates the solution of the autonomous electricity trading problem. Previous research in other trading agent competitions (TAC) used utility-optimization approaches, however the domains, and consequently the methods used were different than LATTE. Other approaches to Power TAC broker design either did not explicitly optimize the actions’ predicted utility, or used tariff optimization strategies that could be viewed as restricted cases of LATTE. The TACTEX agents, which achieved state-of-the-art performance in international competitions and controlled experiments, are a unique contribution of this dissertation. The empirical analysis of the importance of different components of broker agents using LATTE is another unique contribution of this dissertation. Finally, this dissertation is the first to investigate the impact of Time-Of-Use (TOU) tariffs used by autonomous brokers in competitive markets with autonomous customer agents, in a detailed, large-scale, realistic retail-market simulation such as Power TAC.

## Chapter 9

# Conclusions and Future Work

Motivated by the Power TAC domain, and by the potential contribution of autonomous retail broker agents to future smart grids, this dissertation contributes a general algorithm for autonomous trading in modern electricity markets and analyzes its impact on autonomous brokers and on the economy. This chapter reviews the dissertation’s scientific contributions to the areas of artificial intelligence, smart grids, and electricity markets (Section 9.1), and then discusses promising directions for future work in the challenging domain of autonomous electricity trading (Section 9.2).

### 9.1 Contributions

The five main contributions of this dissertation are summarized as follows.

1. The **problem formalization** of autonomous retail broker trading in modern electricity markets presented in Chapter 3 is suitable when (a) a broker trades in the retail market by publishing tariff contracts, (b) a broker trades in the wholesale market by bidding for future contracts in a sequence of auctions, and (c) electricity supply-demand imbalance results in payments to or by the

broker. This problem is a Partially Observable Markov Decision Process, however for computational tractability and modeling clarity, we formalize it as a Markov Decision Process, which due to its complexity is intractable to solve exactly. Therefore, this problem formalization provides a guideline for approximate solutions to the trading problem.

2. **LATTE** (Chapter 4) is a general algorithm for real-time autonomous trading in modern electricity markets. **LATTE**, (Lookahead-policy for Autonomous Time-constrained Trading of Electricity) approximates the solution to the autonomous broker trading problem using a lookahead policy that efficiently samples action combinations and predicts their expected utility over a future horizon. **LATTE** interleaves action sampling with action-effect predictions to constrain the number of action-combinations that it examines. **LATTE** is a general framework that can be instantiated in different ways that tailor it to specific setups.
3. The **TacTex agents** are fully implemented and operational agents that performed successfully at international tournaments and controlled experiments. This dissertation contributes their binaries, as well as their source code and other related resources (see Appendix A). The **TacTex** agents use different instantiations of **LATTE** (Chapters 5, 6, 7) and can serve as benchmarks for future research in the power trading domain.
4. **Extensive empirical analysis** (Chapters 5, 6, 7) validates the effectiveness and robustness of the instantiations of **LATTE** to different competition levels and under a variety of environmental conditions, shedding light on the main reasons for **LATTE**'s success by examining the importance of its constituent components.
5. **The impact of Time-Of-Use tariffs in competitive markets** on an au-

onomous broker and on the economy was analyzed empirically. Time-Of-Use tariffs are a main method proposed for demand-side management both in the literature and in real-markets. One of the instantiations of LATTE contributes a principled, gradient-ascent algorithm for optimizing Time-Of-Use tariffs by an autonomous broker in competitive markets.

Empirical analysis shows that a self-interested broker can use Time-Of-Use tariffs to benefit itself, and by doing so it benefits its customers and the economy. While Time-Of-Use tariffs can induce *customer-herding*, our Time-Of-Use broker prevented it by implicitly coordinating flattening through profit-maximizing tariffs. This coordinated flattening underlines a potential benefit of employing autonomous Time-Of-Use brokers in competitive power markets. In addition, we have seen that a Time-Of-Use broker’s customer share is an important factor in its ability to flatten demand: to counter-balance peaked consumption of fixed-rate brokers’ customers, it needs to gain large customer-share by creating attractive Time-Of-Use tariffs that are still profitable.

## 9.2 Future Work

While the TACTEX agents performed successfully in competitions and controlled experiments using LATTE, there are still many important challenges to be addressed in the autonomous electricity trading domain. This chapter surveys promising directions for future work of extending LATTE within Power TAC (Section 9.2.1), of extending Power TAC to encourage productive future work on real-world problems, (Section 9.2.2), and of extending LATTE towards working in real-world markets (Section 9.2.3).

### 9.2.1 Directions for Extending LATTE within Power TAC

This section surveys promising directions for extending LATTE within the Power TAC domain.

**Using production tariffs for renewable energy:** production tariffs are contracts for procuring energy from renewable producers, and as such they are the main tool for acquiring clean energy resources. While our instantiations of LATTE have not sampled production tariff actions, LATTE includes the framework to use them. The only change that needs to be made is to sample production tariff actions is a new instantiation of the `ComputeNextCandidateTariffAction()` function. Our initial instantiation of LATTE with production tariffs shows promising results. This instantiation executes at any given timeslot either lookahead trajectories with consumption tariff actions, or lookahead trajectories with production tariff actions, interleaving between the two across different timeslots. In this way, the observed effect of one type of tariff action is taken into account in the lookahead with the other type of tariff action, forming an incremental local optimization procedure.

**Optimizing supply-demand imbalance:** While in general a broker should aim to perfectly balance the supply and demand in its portfolio, in some situations a broker can benefit from having an imbalanced portfolio, for instance when such an imbalance contributes to lower the total imbalance of all brokers. Referring back to the `Reward()` function in LATTE (lines 32-34), the  $bal(I_{B_0,t})$  term which determines the imbalance payments can be positive in such cases. By sampling non-zero imbalance levels in the `SampleImbalanceLevels()` function (line 9 of LATTE), and by learning the  $bal(I_{B_0,t})$  function online, a broker could predict the compensation for different imbalance levels and use them in its lookahead inside the `Reward()` function.



**Direct Load Control (DLC):** Direct load control is a demand-side management method that presents an alternative to pricing-based demand-side management. In Power TAC direct load control can be achieved by balancing-market actions that curtail consumption and control storage devices. Extending LATTE to use such actions would equip it with an important class of demand-side management actions. To use such actions, LATTE will have to examine combinations of wholesale-, tariff- and balancing-market actions in its lookahead, which may increase its lookahead complexity by an order of magnitude. Finding an efficient way to incorporate such actions into LATTE’s lookahead is an important research direction. An initial implementation could treat balancing actions hierarchically (similarly to how LATTE treats wholesale actions), sample imbalance levels to aim for (see previous item), and then examine combinations of wholesale-procurement and curtailment quantities that achieve the desired imbalance. The curtailment orders’ limit prices could be optimized at a lower level similarly to LATTE’s wholesale bidding algorithm.

**Real-Time Pricing (RTP) tariffs:** Real-Time Pricing (RTP) tariffs have been proposed as a pricing-based demand-side management method. While LATTE could be extended to sample such tariffs in its lookahead, the decision-making process of setting these prices in real-time is outside the scope of LATTE, and would required extending LATTE’s framework. A possible first step in this direction could be parameterizing and sampling RTP tariffs by their expected prices in `ComputeNextCandidateTariffAction()`, then publishing a tariff based on LATTE’s lookahead, and then setting real-time prices using a lookahead thread that executes between tariff publications. This lookahead thread will need to select a real-time price at each timeslot, and each such selection will affect customers’ consumption in the following timeslots. Even with price discretization, the lookahead search will increase combinatorially. To keep

the complexity of search within practical bounds, Monte-Carlo Tree-Search methods such as UCT [50] may turn out useful.

**Time-Of-Use (TOU) free riders:** Time-Of-Use tariffs can create a free-rider phenomenon, in which brokers who use fixed-rate tariffs and do not contribute to peak-flattening, enjoy the reduced prices caused by Time-Of-Use brokers. An interesting direction for future research is the question of how to incentivize brokers more fairly, such that brokers' contribution to peak-flattening would affect their compensation. A first step in this direction is quantifying the monetary impact of a broker's contribution to peak-flattening and use it to derive a broker's compensation for its contribution.

**Theoretical analysis of Time-Of-Use (TOU) tariffs:** This dissertation has analyzed Time-Of-Use tariffs empirically. An important and interesting direction for future research is to augment the empirical results with a theoretical analysis of the properties and impact of Time-Of-Use tariffs used by autonomous brokers in competitive retail markets. Such an analysis could characterize when TOU tariffs would be beneficial for the broker and for the economy, characterize how a broker's market-share affect TOU tariffs' benefit for the broker and for the economy, and provide theoretical guarantees on the quality of approximate solutions such as LATTE-TOU's.

**Time-Of-Use tariffs in presence of renewable generation:** We investigated the usage of Time-Of-Use tariffs by brokers to reduce the peak-demand. An important potential use for Time-Of-Use tariffs is adapting demand to match the availability of intermittent, renewable resources such as solar and wind. LATTE-TOU has all the required components in place; in our code, all that is needed is to turn on the flag that enables production tariff publications, and investigate LATTE-TOU's behavior in presence of these production tariffs.

**Time-Of-Use tariffs in presence of many competing brokers:** Investigating the impact of Time-Of-Use tariffs on brokers and on the economy in the presence of highly competitive markets is an important direction for future research. Electricity markets are gradually being opened to competition, and Time-Of-Use pricing is one of the main methods proposed for demand-side management. Therefore, the combination of Time-Of-Use tariffs and a highly competitive market scenario that could be of interest to power market designers. An interesting first step in this direction would be running different mixes of TacTex-TOU with FixedRate brokers (see Chapter 7), at different competition levels, e.g. 4-, 6-, 8-broker games. For example, in 4-agent games, it would be interesting to test 4 brokers of the same type, then 3 brokers of one type and 1 broker of the other type, and 2 brokers of each type. It would be interesting to examine whether increased competition solves the problem of free-riders, and whether there are stronger or weaker financial incentives for brokers to use TOU tariffs.

**Contract hedging:** In Power TAC, a broker can hedge itself in the tariff market by using early withdraw penalties and periodic fees. Although not covered in this dissertation, one of our brokers<sup>1</sup> uses a version of LATTE that samples fixed-rate tariffs with fixed early-withdraw fees. A more general implementation would optimize the early-withdraw fees based on risk considerations.

**Bayesian estimation of hidden information:** LATTE uses heuristics and expected values to estimate hidden information, such as the number of subscriptions to competing tariffs. Pardoe et al. used a particle filter for estimating hidden information [75] in the TAC Ad Auctions competition. Such an approach may turn out to be useful for achieving better transition function predictions, i.e. predicting customer responses, future demand, and auction results.

---

<sup>1</sup>TacTex-14 which was not covered in this dissertation, see Chapter 6.

### 9.2.2 Directions for Extending Power TAC

This section surveys promising directions for extending Power TAC to encourage productive future work on real-world problems.

**Tightening the development feedback loop - faster simulations:** To accelerate development progress in Power TAC it would be useful to tighten the development feedback loop, at different levels. At the simulation level, it would be useful to shorten the simulation time from the current time of 2 hours. Having shorter simulations would allow for running more experiments and getting insights more quickly. Shortening the simulation time is a challenging task: out of the 5 seconds allocated per timeslot, the simulator typically uses less than 2 seconds, leaving brokers 3 seconds for computation. To shorten the time allocated per timeslot, the simulator, and potentially brokers, will have to cut computation time significantly. This cut would require either profiling the code and finding opportunities for increased efficiency, or abstracting processes. The former is more desirable, but it is unclear if there is significant room for improvement.

**Tightening the development feedback loop - frequent benchmarking:** The Power TAC competition is an effective way to encourage research progress. Typically, progress is fastest around competition time, due to the opportunities to test broker agents against the most recent versions of competitors (rather than against previous year’s competitors). To accelerate progress, we suggest a setup in which all teams get to test their agent against the most recent, state-of-the-art brokers that would be available to play against. Such a setup could be a continual web-based tournament where everyone can upload binaries and test them against others at any time. Such a setup would accelerate research, development, and could take even better advantage of the

Power TAC simulation.

**Increased realism - customer rationality:** Power TAC customers are modeled as being imperfectly rational, and as having inertia that makes them reluctant to switch tariffs too frequently. These two properties model the effects of imperfect information, and human decision making. It is reasonable to believe that as autonomous customer agents become more widespread, information flow will improve, autonomous customer agents will be able to make more rational decisions, and they will not be inhibited by inertia. Power TAC supports different customer rationality and inertia levels, which are easily configurable. It would be interesting to test the dynamics of markets in presence of more or less rational customers.

Past research in this domain included the work of [66], and some preliminary research that we have done. In our experiments, we observed that with full rationality and no inertia there are large fluctuations in customer subscriptions, since customers always subscribe to the best tariff as soon as it was published. Such phenomena could affect the stability of the electricity grid; there is interesting research to be done in investigating methods to mitigate this effect.

**Increased realism - cost-comfort trade-off:** In Power TAC, customer discomfort is an  $L_2$  norm of the difference between their desired consumption and the actual consumption. In the real-world the discomfort model is more complex, incorporating different hard constraints. Incorporating more complex discomfort functions into Power TAC and into LATTE's models will open an interesting research direction, which will test the generality of LATTE across different discomfort functions, and prepare it to operate in the real-world.

**Increased realism - strong incentives to balance supply and demand:** Strengthening

the financial incentives to balance supply and demand in the simulator would encourage brokers to use balancing actions more frequently, and come up with new balancing strategies. Supply-demand balance is an important issue in the real-world, and becomes more challenging with renewable, intermittent generation. Therefore, developing balancing strategies could have a significant real-world impact.

**Increased realism - running experiments in equilibrium mode:** Broker profits are bounded by the default tariffs proposed to customers by the simulator. Currently, these default tariffs provide room for brokers to cooperate and sell high above their marginal costs. Such cooperation has happened mainly in low competition levels, such as 2-broker games, and is an artifact of Power TAC simulating market liberalization at each game. In the real-world, the more common situation is that prices settle around some equilibrium starting some point in time. We believe that testing games with prices around price equilibriums (by narrowing the margin of the default simulation tariffs) can provide interesting, useful insights on real-world dynamics.

**Increased realism - line capacity limitation:** With larger populations of distributed generation and distributed storage customers, line capacity limitations may come into effect. Modeling line capacity limitations in Power TAC is a challenging task, but could contribute to further advance the simulated dynamics towards the real-world dynamics.

**Increased realism - power factor effects:** A customer's power factor affects energy losses, and could therefore affect brokers' decisions on how to charge customers with high energy losses. Adding power factor effects into Power TAC will increase the realism of the retail market dynamics.

**Increased realism - simulation time resolution:** Power TAC's simulation time

progresses in discrete, 1-hour steps. While this allows to model period of months, it limits the resolution in which electrical and market events can be simulated. Increasing the resolution of the simulation is a challenging task, however if it were to be done, it would further advance the simulator towards real-world dynamics.

### 9.2.3 Directions for Extending LATTE to Real Markets

An important goal for future research is extending LATTE to being usable in real-world markets. The previous two sections laid out two groups of future directions, one for extending LATTE within Power TAC (Section 9.2.1), and the other for further increasing the realism of Power TAC (Section 9.2.2). Both groups of future directions should get LATTE closer to being operational in real-markets, by encouraging extended and refined instantiations of LATTE. This section lists steps that we view as important milestones for deploying LATTE in real-world markets:

**Real-time smart-meter readings:** The Power TAC simulator assumes that all customers have smart-meters that report their readings to the broker once per hour, in real-time. In many real-world regions, smart-meters are not installed yet. Moreover, in many of the regions where smart-meters are installed, readings are not sent in real-time to the retailer, but rather with some delay, e.g. once per day and delayed by 48 hours. The ability of a broker to observe customer readings in real-time is important for acting and responding in real-time. While LATTE could still work with delayed readings, it would work more accurately with real-time readings.

**Autonomous customer agents:** In Power TAC, autonomous agents optimize customers' cost and comfort by subscribing to tariffs and adapting customers' consumption. These autonomous agents are more efficient and reactive than human customers. To get the most benefit out of LATTE, it would be useful to

use it in retail markets where such autonomous customer agents are installed. Therefore, deploying autonomous customer agents in buildings can facilitate the testing of LATTE in the real-world. LATTE does not inherently depend on the existence of such agents, however their increased responsiveness will give LATTE a larger space of actions to optimize over.

**Small field tests:** Following up on the previous item, an important step towards deploying LATTE in the real-world would be to test it in small field tests, with tens, then hundreds of homes that employ autonomous customer agents. Since the resulting environment would include tens or hundreds of agents, such tests may be able to reveal potential destructive combinations of agent behaviors, and refine/constrain both LATTE and the autonomous customers behaviors.

**Worst-case behavior:** While in Power TAC the goal of a broker is to maximize its *expected* profit, in the real-world customer agreement to participate in automated markets would probably also depend on their level of exposure to worst-case events. Therefore, an important direction for future research is extending LATTE to incorporate worst-case and risk considerations.

**Learning predictors from data:** Our empirical analysis revealed the importance of having accurate demand-predictions to LATTE’s performance. In the real-world, customer subscription behavior, consumption elasticity, and consumption shifting may have complex functional patterns. These behaviors will have to be learned effectively from real-world data. Similarly, cost-predictions will have to be learned from real-world data. Learning predictors from real-world data will likely require new instantiations of LATTE, with new implementations of the `PredictTariffEffects()` (line 7 of LATTE) and `PredictWholesalePrice()` (line 16 of LATTE) functions.



### 9.3 Concluding Remarks

This dissertation lays a foundation for understanding how autonomous electricity trading broker agents should operate in real-time modern electricity markets, by contributing a general decision-making framework for such brokers. Based on this foundation, effective and robust autonomous brokers can be designed. Such brokers could be used as a basis for research on how to design electricity markets that would utilize such brokers most effectively to the benefit of clean, reliable, and sustainable energy. I believe that AI will increasingly be used for the benefit of the smart grid, and I hope that this dissertation will contribute to a better understanding of how to build intelligent agents that will take part in addressing the complex challenges faced by future energy delivery systems.

# Appendix A

## TacTex Agents Source Code, Binaries, and Resources

The source code and binaries of the TACTEX agents, as well as related resources are an online appendix. The online appendix additionally includes links to official simulator versions, as well as source code of specific versions used for experimentation in this dissertation. The online appendix is located in the following URL:

<http://www.cs.utexas.edu/~urieli/thesis>

## Appendix B

# Power TAC Game Parameters

Table B.1: Parameters used in Power TAC tournament games. Source: The Power TAC game specification [46].

Parameter	Standard Game Setting
Length of pre-game bootstrap period	14 days
Nominal length of game	60 days
Probability of game end for each time slot after time slot 1320 (start of day 55)	$\frac{1}{121}$
Minimum game length	1320
Expected game length	1440
Timeslot length	60 minutes
Time compression ratio	720 (5 seconds/time slot)
Open time slots on wholesale market	24
Market closing time	1 time slot ahead
Minimum order quantity	0.1 kWh
Distribution fee	[0.003 - 0.03]€/kWh
Balancing price basis	most recent clearing price
Balancing cost	[0.02 - 0.06]€/kWh
Slope of regulating market price	$10^{-6}$ , $10^{-6}$ €/kWh
Default broker's min and max bid order prices	-100, -5
Default broker's min and max ask order prices	0.1, 30
Tariff publication fee	[1000 - 5000] €
Tariff revocation fee	[100 - 500] €
Tariff publication interval	6 time slots
Daily bank debt interest rate	4.0%/365 $\dots$ 12.0%/365
Daily bank deposit interest rate	0.5 $\beta$
Weather report interval	1 hour
Weather forecast interval	1 hour
Weather forecast horizon	24 hours

# Bibliography

- [1] M. Albadi and E. El-Saadany. A summary of demand response in electricity markets. *Electric Power Systems Research*, 78(11):1989 – 1996, 2008.
- [2] S. Amin and B. Wollenberg. Toward a smart grid: power delivery for the 21st century. *Power and Energy Magazine, IEEE*, 3(5):34–41, Sept 2005.
- [3] C. G. Atkeson, A. W. Moore, and S. Schaal. Locally weighted learning. 1997.
- [4] J. Babic and V. Podobnik. Adaptive bidding for electricity wholesale markets in a smart grid. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2014)*, May 2014.
- [5] J. Babic and V. Podobnik. An analysis of Power Trading Agent Competition 2014. In S. Ceppi, E. David, V. Podobnik, V. Robu, O. Shehory, S. Stein, and I. A. Vetsikas, editors, *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, volume 187 of *Lecture Notes in Business Information Processing*, pages 1–15. Springer International Publishing, 2014.
- [6] D. Bertsekas and D. Castanon. Rollout algorithms for stochastic scheduling problems. *Journal of Heuristics*, 5(1):89–108, 1999.

- [7] S. Borenstein. The trouble with electricity markets: Understanding California’s restructuring disaster. *Journal of Economic Perspectives*, 16(1):191–211, 2002.
- [8] D. Callaway and I. Hiskens. Achieving controllability of electric loads. *Proceedings of the IEEE*, 99(1):184–199, Jan 2011.
- [9] J. Casazza, J. Casazza, and F. Delea. *Understanding electric power systems: an overview of the technology and the marketplace*, volume 13. John Wiley & Sons, 2003.
- [10] E. Celebi and J. Fuller. A model for efficient consumer pricing schemes in electricity markets. *Power Systems, IEEE Transactions on*, 22(1):60–67, Feb 2007.
- [11] C. Chen, S. Kishore, and L. Snyder. An innovative RTP-based residential power scheduling scheme for smart grids. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 5956–5959, May 2011.
- [12] M. Chick. Le tarif vert retrouve: The marginal cost concept and the pricing of electricity in britain and france, 1945-1970. *The Energy Journal*, pages 97–116, 2002.
- [13] G. Conzelmann, G. Boyd, V. Koritarov, and T. Veselka. Multi-agent power market simulation using EMCAS. In *Power Engineering Society General Meeting, 2005. IEEE*, pages 2829–2834. IEEE, 2005.
- [14] S. Datchanamoorthy, S. Kumar, Y. Ozturk, and G. Lee. Optimal time-of-use pricing for residential load control. In *Smart Grid Communications (Smart-GridComm), 2011 IEEE International Conference on*, pages 375–380, Oct 2011.

- [15] A. David and Y. Li. Consumer rationality assumptions in the real time pricing of electricity. In *Advances in Power System Control, Operation and Management, 1991. APSCOM-91., 1991 International Conference on*, pages 391–396 vol.1, Nov 1991.
- [16] A. David and Y. Li. Effect of inter-temporal factors on the real time pricing of electricity. *Power Systems, IEEE Transactions on*, 8(1):44–52, Feb 1993.
- [17] Electric Power Research Institute. *Estimating the Costs and Benefits of the Smart Grid*, 2011.
- [18] European Commission. *Communication: Energy Roadmap 2050*, 2011.
- [19] P. Faria and Z. Vale. Demand response in electrical energy supply: An optimal real time pricing approach. *Energy*, 36(8):5374 – 5384, 2011.
- [20] H. Finnsson and Y. Björnsson. Simulation-based approach to general game playing. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 1, AAAI’08*, pages 259–264. AAAI Press, 2008.
- [21] V. K. Garg, T. Jayram, and B. Narayanaswamy. Online optimization with dynamic temporal uncertainty: Incorporating short term predictions for renewable integration in intelligent energy systems. In *AAAI*, 2013.
- [22] S. Gelly and D. Silver. Combining online and offline knowledge in UCT. In *Proceedings of the 24th International Conference on Machine Learning, ICML ’07*, pages 273–280, New York, NY, USA, 2007. ACM.
- [23] C. P. Gomes. Computational sustainability: Computational methods for a sustainable environment, economy, and society. *The Bridge*, 39(4):5–13, 2009.
- [24] A. Greenwald. The 2002 trading agent competition: An overview of agent strategies. *AI Magazine*, 24(1):83, 2003.

- [25] A. R. Greenwald and J. O. Kephart. Shopbots and pricebots. In *Agent Mediated Electronic Commerce II*, pages 1–23. Springer, 2000.
- [26] A. R. Greenwald and J. O. Kephart. Probabilistic pricebots. In *Proceedings of the fifth international conference on Autonomous agents*, pages 560–567. ACM, 2001.
- [27] A. R. Greenwald, J. O. Kephart, and G. J. Tesauro. Strategic pricebot dynamics. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 58–67. ACM, 1999.
- [28] G. Grozev, D. Batten, M. Anderson, G. Lewis, J. Mo, and J. Katzfey. NEM-SIM: Agent-based simulator for australia’s national electricity market. In *SimTecT 2005 Conference Proceedings, Sydney, Australia*. Citeseer, 2005.
- [29] M. He, A. Rogers, E. David, and N. R. Jennings. Designing and evaluating an adaptive trading agent for supply chain management applications. In H. L. Poutre, N. Sadeh, and J. Sverker, editors, *Agent-mediated Electronic Commerce, Designing Trading Agents and Mechanisms: AAMAS 2005 Workshop AMEC 2005, Utrecht, Netherlands, July 25, 2005, and IJCAI 2005 Workshop TADA 2005, Edinburgh, UK, August 1, 2005, Selected and Revised Papers*, pages 35–42. Springer, 2005. Event Dates: August 2005.
- [30] S. P. Holland and E. T. Mansur. The distributional and environmental effects of time-varying prices in competitive electricity markets. *UC Energy Institute CSEM Working Paper*, (143), 2005.
- [31] J. Hoogland and H. L. Poutre. An effective broker for the Power TAC 2014. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2015)*, May 2015.

- [32] I. Horowitz and C. Woo. Designing pareto-superior demand-response rate options. *Energy*, 31(6):1040–1051, 2006.
- [33] P. R. Jordan and M. P. Wellman. Designing an ad auctions game for the trading agent competition. In *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pages 147–162. Springer, 2010.
- [34] P. Joskow and J. Tirole. Retail electricity competition. *The RAND Journal of Economics*, 37(4):799–815, 2006.
- [35] P. L. Joskow. Lessons learned from electricity market liberalization. *The Energy Journal*, Volume 29, 2008.
- [36] M. Kahlen and W. Ketter. Aggregating electric cars to sustainable virtual power plants: The value of flexibility in future electricity markets. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15)*, Jan. 2015.
- [37] M. Kahlen, W. Ketter, and J. van Dalen. Balancing with electric vehicles: A profitable business model. 2014.
- [38] E. C. Kara, Z. Kolter, M. Berges, B. Krogh, G. Hug, and T. Yuksel. A moving horizon state estimator in the control of thermostatically controlled loads for demand response. In *Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on*, pages 253–258. IEEE, 2013.
- [39] M. J. Kearns, Y. Mansour, and A. Y. Ng. Approximate planning in large pomdps via reusable trajectories. In *Advances in Neural Information Processing Systems 12, [NIPS Conference, Denver, Colorado, USA, November 29 - December 4, 1999]*, pages 1001–1007, 1999.



- [40] M. Kefayati and R. Baldick. On optimal operation of storage devices under stochastic market prices. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 7576–7581, Dec 2013.
- [41] M. Kefayati and R. Baldick. Anticipative charging of plug-in electric vehicles and its impact on the grid. In *Transportation Electrification Conference and Expo (ITEC), 2014 IEEE*, pages 1–6, June 2014.
- [42] J. O. Kephart and A. R. Greenwald. Shopbot economics. In *Game theory and decision theory in agent-based systems*, pages 119–158. Springer, 2002.
- [43] J. O. Kephart, J. E. Hanson, and A. R. Greenwald. Dynamic pricing by software agents. *Computer Networks*, 32(6):731–752, 2000.
- [44] W. Ketter, J. Collins, and C. A. Block. Smart grid economics: Policy guidance through competitive simulation. 2010.
- [45] W. Ketter, J. Collins, and P. Reddy. Power TAC: A competitive economic simulation of the smart grid. *Energy Economics*, 39(0):262 – 270, 2013.
- [46] W. Ketter, J. Collins, P. P. Reddy, and M. D. Weerdt. The 2015 Power Trading Agent Competition. *ERIM Report Series Reference No. ERS-2015-001-LIS*, 2015.
- [47] W. Ketter, M. Peters, and J. Collins. Autonomous agents in future energy markets: The 2012 Power Trading Agent Competition. In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*. AAAI, 2013.
- [48] C. Kiekintveld, Y. Vorobeychik, and M. Wellman. An analysis of the 2004 supply chain management trading agent competition. In H. Poutr, N. Sadeh, and S. Janson, editors, *Agent-Mediated Electronic Commerce. Designing Trading Agents and Mechanisms*, volume 3937 of *Lecture Notes in Computer Science*, pages 99–112. Springer Berlin Heidelberg, 2006.

- [49] D. Kirschen. Demand-side view of electricity markets. *Power Systems, IEEE Transactions on*, 18(2):520–527, May 2003.
- [50] L. Kocsis and C. Szepesvári. Bandit based Monte-Carlo planning. In *Machine Learning: ECML 2006*, pages 282–293. Springer, 2006.
- [51] N. Kohl and P. Stone. Machine learning for fast quadrupedal locomotion. In *The Nineteenth National Conference on Artificial Intelligence*, pages 611–616, July 2004.
- [52] J. Z. Kolter and J. Ferreira Jr. A large-scale study on predicting and contextualizing building energy usage. 2011.
- [53] J. Z. Kolter and T. Jaakkola. Approximate inference in additive factorial hmms with application to energy disaggregation. In *International conference on artificial intelligence and statistics*, pages 1472–1482, 2012.
- [54] J. Z. Kolter, Z. Jackowski, and R. Tedrake. Design, analysis, and learning control of a fully actuated micro wind turbine. In *American Control Conference (ACC), 2012*, pages 2256–2263. IEEE, 2012.
- [55] J. Z. Kolter and M. J. Johnson. Redd: A public data set for energy disaggregation research. In *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, volume 25, pages 59–62. Citeseer, 2011.
- [56] R. T. Kuate, M. Chli, and H. H. Wang. Optimising market share and profit margin: SMDP-based tariff pricing under the smart grid paradigm. In *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*, pages 1–6, Oct 2014.
- [57] R. T. Kuate, M. He, M. Chli, and H. H. Wang. An intelligent broker agent for energy trading: An MDP approach. In *The 23rd International Joint Conference on Artificial Intelligence*, 2013.

- [58] J.-y. Kwak, D. Kar, W. B. Haskell, P. Varakantham, and M. Tambe. Building thinc: user incentivization and meeting rescheduling for energy savings. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 925–932. International Foundation for Autonomous Agents and Multiagent Systems, 2014.
- [59] D. Lee and R. Baldick. Analyzing the variability of wind power output through the power spectral density. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–8. IEEE, 2012.
- [60] D. Lee and R. Baldick. Future wind power scenario synthesis through power spectral density analysis. *Smart Grid, IEEE Transactions on*, 5(1):490–500, Jan 2014.
- [61] D. Lee and R. Baldick. Short-term wind power ensemble prediction based on Gaussian processes and neural networks. *Smart Grid, IEEE Transactions on*, 5(1):501–510, 2014.
- [62] H. Li and L. Tesfatsion. Development of open source software for power market research: The AMES test bed. *Journal of Energy Markets*, 2(2):111–128, 2009.
- [63] B. Liefers, J. Hoogland, and H. L. Poutre. A successful broker agent for Power TAC. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2014)*, May 2014.
- [64] R. Lorentz. Amazons discover Monte-Carlo. In H. van den Herik, X. Xu, Z. Ma, and M. Winands, editors, *Computers and Games*, volume 5131 of *Lecture Notes in Computer Science*, pages 13–24. Springer Berlin Heidelberg, 2008.
- [65] A.-H. Mohsenian-Rad and A. Leon-Garcia. Optimal residential load control

- with price prediction in real-time electricity pricing environments. *Smart Grid, IEEE Transactions on*, 1(2):120–133, Sept 2010.
- [66] E. Ntagka, A. Chrysopoulos, and P. A. Mitkas. Designing tariffs in a competitive energy market using particle swarm optimization techniques. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2014)*, May 2014.
- [67] S. Ozdemir and R. Unland. AgentUDE: The success story of the Power TAC 2014’s champion. In *AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2015)*, May 2015.
- [68] P. Palensky and D. Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *Industrial Informatics, IEEE Transactions on*, 7(3):381–388, Aug 2011.
- [69] A. A. Panagopoulos, M. Alam, A. Rogers, and N. Jennings. AdaHeat: A general adaptive intelligent agent for domestic heating control. In *14th International Conference on Autonomous Agents and Multi-Agent Systems*, 2015.
- [70] A. A. Panagopoulos, G. Chalkiadakis, and R. N. Jennings. Towards optimal solar tracking: a dynamic programming approach. In *AAAI-2015: 29th AAAI Conference on Artificial Intelligence*, pages 695–701, 2015.
- [71] D. Pardoe, D. Chakraborty, and P. Stone. TacTex09: A champion bidding agent for ad auctions. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, May 2010.
- [72] D. Pardoe and P. Stone. Predictive planning for supply chain management. In *Proceedings of the International Conference on Automated Planning and Scheduling*, June 2006.

- [73] D. Pardoe and P. Stone. TacTex-2005: A champion supply chain management agent. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pages 1489–94, July 2006.
- [74] D. Pardoe and P. Stone. Designing adaptive trading agents. *ACM SIGecom Exchanges*, 10(2):37–9, June 2011.
- [75] D. Pardoe and P. Stone. A particle filter for bid estimation in ad auctions with periodic ranking observations. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, May 2011.
- [76] D. M. Pardoe. *Adaptive Trading Agent Strategies Using Market Experience*. PhD thesis, 2011.
- [77] M. Peters, W. Ketter, M. Saar-Tsechansky, and J. Collins. A reinforcement learning approach to autonomous decision-making in smart electricity markets. *Machine Learning*, 92(1):5–39, 2013.
- [78] M. Pipattanasomporn, H. Feroze, and S. Rahman. Multi-agent systems in a distributed smart grid: Design and implementation. In *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*, pages 1–8. IEEE, 2009.
- [79] W. Powell and S. Meisel. Tutorial on stochastic optimization in energy – part ii: An energy storage illustration. *Power Systems, IEEE Transactions on*, PP(99):1–8, 2015.
- [80] W. B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality, 2nd Edition*. Wiley, 2011.
- [81] I. Praça, C. Ramos, Z. Vale, and M. Cordeiro. MASCEM: a multiagent system

- that simulates competitive electricity markets. *Intelligent Systems, IEEE*, 18(6):54–60, 2003.
- [82] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994.
- [83] S. Ramchurn, P. Vytelingum, A. Rogers, and N. Jennings. Agent-based control for decentralised demand side management in the smart grid. In *The Tenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, pages 5–12, 2011.
- [84] S. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings. Putting the “smarts” into the smart grid: A grand challenge for artificial intelligence. *Communications of the ACM*, 55(4):86–97, 2012.
- [85] P. P. Reddy and M. M. Veloso. Learned behaviors of multiple autonomous agents in smart grid markets. In *AAAI*, 2011.
- [86] P. P. Reddy and M. M. Veloso. Strategy learning for autonomous agents in smart grid markets. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Two*, pages 1446–1451. AAAI Press, 2011.
- [87] P. P. Reddy and M. M. Veloso. Factored Models for Multiscale Decision Making in Smart Grid Customers. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI-12)*, 2012.
- [88] P. P. Reddy and M. M. Veloso. Negotiated Learning for Smart Grid Agents: Entity Selection based on Dynamic Partially Observable Features. In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI-13)*, 2013.

- [89] A. Rogers, S. Ghosh, R. Wilcock, and N. R. Jennings. A scalable low-cost solution to provide personalised home heating advice to households. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, BuildSys'13, pages 1:1–1:8, New York, NY, USA, 2013. ACM.
- [90] A. Rogers, S. Maleki, S. Ghosh, and J. Nicholas R. Adaptive home heating control through Gaussian process prediction and mathematical programming. In *Second International Workshop on Agent Technology for Energy Systems (ATES 2011)*, pages 71–78, May 2011.
- [91] S. Roth. Major changes coming to California electricity rates. *The Desert Sun*, May 2015.
- [92] G. A. Rummery and M. Niranjan. On-line Q-learning using connectionist systems. 1994.
- [93] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition, 2009.
- [94] N. Sadeh, R. Arunachalam, J. Eriksson, N. Finne, and S. Janson. TAC-03—a supply-chain trading competition. *AI magazine*, 24(1):92, 2003.
- [95] P. Samadi, A.-H. Mohsenian-Rad, R. Schober, V. Wong, and J. Jatskevich. Optimal real-time pricing algorithm based on utility maximization for smart grid. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 415–420, Oct 2010.
- [96] F. Schweppe, B. Daryanian, and R. Tabors. Algorithms for a spot price responding residential load controller. *Power Systems, IEEE Transactions on*, 4(2):507–516, May 1989.
- [97] D. Silver and J. Veness. Monte-Carlo planning in large POMDPs. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Ad-*

- vances in Neural Information Processing Systems 23*, pages 2164–2172. Curran Associates, Inc., 2010.
- [98] K. Spees and L. Lave. Impacts of responsive load in PJM: load shifting and real time pricing. *The Energy Journal*, pages 101–121, 2008.
  - [99] S. Stoft. *Power System Economics: Designing Markets for Electricity*. Wiley-IEEE Press, 2002.
  - [100] P. Stone and A. Greenwald. The first international trading agent competition: Autonomous bidding agents. *Electronic Commerce Research*, 5(2):229–65, April 2005.
  - [101] P. Stone, M. L. Littman, S. Singh, and M. Kearns. ATTac-2000: An adaptive autonomous bidding agent. *Journal of Artificial Intelligence Research*, 15:189–206, June 2001.
  - [102] P. Stone, R. E. Schapire, M. L. Littman, J. A. Csirik, and D. McAllester. Decision-theoretic bidding based on learned density models in simultaneous, interacting auctions. *Journal of Artificial Intelligence Research*, 19:209–242, 2003.
  - [103] R. S. Sutton, D. Precup, and S. Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1):181–211, 1999.
  - [104] B. Tao, F. Wu, and G. Chen. Tac AdX’14: Autonomous agents for real-time ad exchange. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, AAMAS ’15, pages 1111–1119, Richland, SC, 2015. International Foundation for Autonomous Agents and Multiagent Systems.



- [105] G. Tesauro and J. L. Bredin. Strategic sequential bidding in auctions using dynamic programming. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 2*, AAMAS '02, pages 591–598, New York, NY, USA, 2002. ACM.
- [106] L. Tesfatsion. Agent-based computational economics: Growing economies from the bottom up. *Artificial life*, 8(1):55–82, 2002.
- [107] C. Triki and A. Violi. Dynamic pricing of electricity in retail markets. *4OR*, 7(1):21–36, 2009.
- [108] United States Department of Energy. *2014 Smart Grid System Report*, August 2014.
- [109] D. Urieli and P. Stone. A learning agent for heat-pump thermostat control. In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, May 2013.
- [110] D. Urieli and P. Stone. Model-selection for non-parametric function approximation in continuous control problems: A case study in a smart energy system. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML'13)*, Sep 2013.
- [111] D. Urieli and P. Stone. TacTex'13: A champion adaptive power trading agent. In *Proceedings of the Twenty-Eighth Conference on Artificial Intelligence (AAAI 2014)*, July 2014.
- [112] D. Urieli and P. Stone. An MDP-Based Winning Approach to Autonomous Power Trading: Formalization and Empirical Analysis. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, May 2016.

- [113] D. Urieli and P. Stone. Autonomous electricity trading using time-of-use tariffs in a competitive market. In *Proceedings of the Thirtieth Conference on Artificial Intelligence (AAAI 2016)*, Feb 2016.
- [114] U.S. Department of Energy. “*Grid 2030*” *A National Vision For Electricity’s Second 100 Years*, 2003.
- [115] U.S.E.I Administration. *Annual Energy Review 2010*, Oct 2010.
- [116] K. Valogianni, W. Ketter, and J. Collins. A multiagent approach to variable-rate electric vehicle charging coordination. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS ’15*, pages 1131–1139, Richland, SC, 2015. International Foundation for Autonomous Agents and Multiagent Systems.
- [117] K. Valogianni, W. Ketter, J. Collins, and D. Zhdanov. Effective management of electric vehicle storage using smart charging. In *Proceedings of 28th AAAI Conference on Artificial Intelligence*, pages 472–478, 2014.
- [118] A. Veit, Y. Xu, R. Zheng, N. Chakraborty, and K. Sycara. Demand side energy management via multiagent coordination in consumer cooperatives. *Journal of Artificial Intelligence Research*, 50:885–922, 2014.
- [119] P. Vytelingum, S. Ramchurn, T. Voice, A. Rogers, and N. Jennings. Agent-based modeling of smart-grid market operations. In *Power and Energy Society General Meeting, 2011 IEEE*, pages 1–8, July 2011.
- [120] P. Vytelingum, S. D. Ramchurn, T. D. Voice, A. Rogers, and N. R. Jennings. Trading agents for the smart electricity grid. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1 - Volume 1*, AAMAS ’10, pages 897–904, Richland, SC, 2010. International Foundation for Autonomous Agents and Multiagent Systems.

- [121] P. Vytelingum, T. D. Voice, S. D. Ramchurn, A. Rogers, and N. R. Jennings. Agent-based micro-storage management for the smart grid. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 39–46. International Foundation for Autonomous Agents and Multiagent Systems, 2010.
- [122] P. Vytelingum, T. D. Voice, S. D. Ramchurn, A. Rogers, and N. R. Jennings. Theoretical and practical foundations of large-scale agent-based micro-storage in the smart grid. *Journal of Artificial Intelligence Research*, pages 765–813, 2011.
- [123] P. Vytelingum, T. D. Voice, S. D. Ramchurn, A. Rogers, and N. R. Jennings. Theoretical and practical foundations of large-scale agent-based micro-storage in the smart grid. *J. Artif. Int. Res.*, 42(1):765–813, Sept. 2011.
- [124] C. J. Watkins and P. Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.
- [125] A. Weidlich and D. Veit. A critical survey of agent-based wholesale electricity market models. *Energy Economics*, 30(4):1728 – 1759, 2008.
- [126] M. P. Wellman, A. Greenwald, and P. Stone. *Autonomous Bidding Agents: Strategies and Lessons from the Trading Agent Competition*. MIT Press, 2007.
- [127] Q. Wu, L. Wang, and H. Cheng. Research of TOU power price based on multi-objective optimization of DSM and costs of power consumers. In *Electric Utility Deregulation, Restructuring and Power Technologies, 2004. (DRPT 2004). Proceedings of the 2004 IEEE International Conference on*, volume 1, pages 343–348 Vol.1, April 2004.
- [128] M. Wytock and J. Z. Kolter. Large-scale probabilistic forecasting in energy systems using sparse gaussian conditional random fields. In *Decision and*

- Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 1019–1024. IEEE, 2013.
- [129] M. Wytock and Z. Kolter. Sparse Gaussian conditional random fields: Algorithms, theory, and application to energy forecasting. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 1265–1273, 2013.
  - [130] P. Yang, G. Tang, and A. Nehorai. A game-theoretic approach for optimal time-of-use electricity pricing. *Power Systems, IEEE Transactions on*, 28(2):884–892, May 2013.
  - [131] S. Yousefi, M. P. Moghaddam, and V. J. Majd. Optimal real time pricing in an agent-based retail market using a comprehensive demand response model. *Energy*, 36(9):5716 – 5727, 2011.
  - [132] Z. Zhou, W. Chan, and J. Chow. Agent-based simulation of electricity markets: a survey of tools. *Artificial Intelligence Review*, 28(4):305–342, 2007.